Computational Fluid- and Solid-Mechanics

Thomas Richter
thomas.richter@ovgu.de
Universität Magdeburg

Carmen Gräßle
graessle@mpi-magdeburg.mpg.de
MPI für Dynamik komplexer technischer Systeme Magdeburg

October 17, 2023
(Version: Sommer 2020)

Contents

Bibliography

[1] H.W. Alt. *Lineare Funktionalanalysis. Eine anwendungsorientierte Einführung*. Springer Verlag, Berlin, 5. auflage edition, 2008.

[2] T. Apel. *Anisotropic finite elements: Local estimates and applications*. Advances in Numerical Mathematics. Teubner, Stuttgart, 1999.

[3] R. Becker and M. Braack. A finite element pressure gradient stabilization for the Stokes equations based on local projections. *Calcolo*, 38(4):173–199, 2001.

[4] M. Braack and P.B. Mucha. Directional do-nothing condition for the Navier-Stokes equations. *J. Comp. Math.*, 32(5):507–521, 2014.

[5] F. Brezzi and M. Fortin. *Mixed and Hybrid Finite Element Methods*. Springer-Verlag, Berlin, 1991.

[6] J. Carlson, A. Jaffe, and A. Wiles, editors. *Millenium Prize Problems*. CMI/ American Mathematical Society, 2006. ISBN-13: 978-0-8218-3679-8.

[7] P.G. Ciarlet. *Mathematical Elasticity, Vol. I: Three-dimensional Elasticity*. North-Holland, Amsterdam., 1991.

[8] P.G. Ciarlet. On Korn's Inequality. *Chin. Ann. Math.*, 31:607–618, 2010.

[9] P. Clément. Approximation by finite element functions using local regularization. *R.A.I.R.O. Anal. Numer.*, 9:77–84, 1975.

[10] D. Coutand and S. Shkoller. Motion of an elastic solid inside an incompressible viscous fluid. *Arch. Ration. Mech. Anal.*, 176:25–102, 2005.

[11] M. Dobrowolski. On the LBB constant on stretched domains. *Math. Nachr.*, 254-255:64–67, 2003.

[12] R.L. Fosdick and E.G. Virga. A variational proof of the stress theorem of Cauchy. *Archive for Rational Mechanics and Analysis*, 105(2):95–103, 1989.

[13] G. P. Galdi. *An Introduction to the Mathematical Theory of the Navier-Stokes Equations, Vol. I, Steady-State Problems*. Springer, New York, 2011.

[14] F. Gazzola and M. Squassina. Global solutions and finite time blow up for damped semilinear wave equations. *Ann. I.H. Poincaré*, 23:185–207, 2006.

[15] V. Girault and P.-A. Raviart. *Finite Elements for the Navier Stokes Equations*. Springer, 1986.

[16] J. Heywood and R. Rannacher. Finite element approximation of the nonstationary Navier-Stokes problem. i. regularity of solutions and second order error estimation for spatial discretizations. *SIAM J. Numer. Anal.*, 19(2):275–311, 1982.

[17] J. Heywood and R. Rannacher. Finite element approximation of the nonstationary Navier-Stokes problem. ii. stability of solutions and error estimates uniform in time. *SIAM J. Numer. Anal.*, 23(4):750–777, 1986.

[18] J. Heywood and R. Rannacher. Finite element approximation of the nonstationary Navier-Stokes problem. iii. smoothing property and higher order error estimates for spatial discretization. *SIAM J. Numer. Anal.*, 25(3):489–512, 1988.

[19] J. Heywood and R. Rannacher. Finite element approximation of the nonstationary Navier-Stokes problem. iv. error analysis for second-order time discretization. *SIAM J. Numer. Anal.*, 27(3):353–384, 1990.

[20] J.G. Heywood, R. Rannacher, and S. Turek. Artificial boundaries and flux and pressure conditions for the incompressible Navier-Stokes equations. *Int. J. Numer. Math. Fluids.*, 22:325–352, 1992.

[21] G.A. Holzapfel. *Nonlinear Solid Mechanics: A Continuum Approach for Engineering*. Wiley-Blackwell, 2000.

[22] C.O. Horgan. Korn's inequalities and their applications in continuum mechanics. *SIAM Review*, 37:491–511, 1995.

[23] T.J.R. Hughes, L.P. Franca, and M. Balestra. A new finite element formulation for computational fluid dynamics: V. circumvent the Babuska-Brezzi condition: A stable Petrov-Galerkin formulation for the Stokes problem accommodating equal order interpolation. *Comput. Methods Appl. Mech. Engrg.*, 59:89–99, 1986.

[24] G.W. Jones and S.J. Chapman. Modeling growth in biological materials. *SIAM Review*, 54(1):52–118, 2012.

[25] A. Linke, G. Matthies, and L. Tobiska. Robust arbitrary order mixed finite element methods for the incompressible stokes equations with pressure independent velocity errors. *ESAIM: Math. Mod. and Num. Anal.*, 50(1):289–309, 2015.

[26] A. Linke, C. Merdon, and W. Wollner. Optimal L^2 velocity error estimate for a modified pressure-robust Crouzeix-Raviart Stokes element. *IMA J. Numer. Anal.*, 37(1):354–374, 2017.

[27] K.-A. Mardal, J. Schöberl, and R. Winther. A uniformly stable Fortin operator for the Taylor-Hood element. *Numer. Math.*, 123(3):537–551, 2012.

[28] M. Mitrea and S. Monniaux. Maximal regularity for the Lamé system in certain classes of non-smooth domains. *Journal of Evolution Equations*, 10(4):811–833, 2010.

[29] R. Rannacher. *Numerik partieller Differentialgleichungen*. Universität Heidelberg, http://numerik.iwr.uni-heidelberg.de/~lehre/notes/, 2008. Vorlesungsskriptum.

[30] R. Rannacher. *Numerik 3. Probleme der Kontinuumsmechanik und ihre numerische Behandlung*. Lecture Notes in Mathematics. Heidelberg University Publishing, 2017. https://doi.org/10.17885/heiup.312.424.

[31] T. Richter. *Fluid-structure Interactions. Models, Analysis and Finite Elements*, volume 118 of *Lecture notes in computational science and engineering*. Springer, 2017.

[32] R.S. Rivlin and J.L. Ericksen. Stress-deformation relations for isotropic materials. *J. of Rational Mech. and Anal.*, 4:323–425, 1955. Reprinted in Rational Mechanics of Materials. International Science Review Series, New York: Gordon & Breach, 1965.

[33] W. Rudin. *Functional Analysis*. International Series in Pure and Applied Mathematics. McGraw-Hill, 1991.

[34] B. Schweizer. *Partielle Differentialgleichungen. Eine anwendungsorientierte Einführung*. Springer, 2013.

[35] L.R. Scott and S. Zhang. Finite element interpolation of nonsmooth functions satisfying boundary conditions. *Math. Comp.*, 54(190):483–493, 1990.

[36] P. Shi and S. Wright. $W^{2,p}$-regularity of the displacement problem for the lamé system on $W^{2,s}$ domains. *J. Math. Anal. Appl.*, 239(2):291 – 305, 1999.

[37] H. Sohr. *The Navier-Stokes Equations. An Elementary Functional Analytic Approach*. Birkhäuser Advanced Texts. Birkhäuser, 2001.

[38] R. Temam. *Navier-Stokes Equations: Theory and Numerical Analysis*. American Mathematical Society, 2000.

[39] C. Truesdell and W. Noll. *The Non-Linear Flied Theories of Mechanics*. Springer-Verlag, 2004.

[40] S. Turek, J. Hron, M. Madlik, M. Razzaq, H. Wobker, and J. Acker. Numerical simulation and benchmarking of a monolithic multigrid solver for fluid–structure interaction problems with application to hemodynamics. Technical report, Fakultät für Mathematik, TU Dortmund, February 2010. Ergebnisberichte des Instituts für Angewandte Mathematik, Nummer 403.

[41] S. Turek, J. Hron, M. Razzaq, H. Wobker, and M. Schäfer. Numerical benchmarking of fluid-structure interaction: A comparison of different discretization and solution approaches. In H.J. Bungartz, M. Mehl, and M. Schäfer, editors, *Fluid Structure Interaction II: Modeling, Simulation and Optimization*. Springer, 2010.

[42] M. von Laer. Finite elemente simulation of non-newtonian flows. Master's thesis, Universität Heidelberg, 2013.

[43] D. Werner. *Funktionalanalysis*. Springer, Berlin, 6. auflage edition, 2007.

Index

List of symbols

δ_{ij}	Kronecker symbol $\delta_{ij} = 1$ if $i = j$ and $\delta_{ij} = 0$ for $i \neq j$
Δ	Laplace- or vector Laplace operator $\Delta = \partial_x^2 + \partial_y^2 + \partial_z^2$
∇	Gradient of scalar or vector gradient operator
$\nabla\cdot$	Divergence
Ω	Domain, usually in $\mathbb{R}^2$ or $\mathbb{R}^3$. Usually Ω has at least Lipschitz regularity
$\partial\Omega$	The boundary of Ω
Ω_h	Triangulation of the domain (finite element mesh)
$K \in \Omega_h$	An element of the triangulation (e.g. triangle, quad, hexahedra, ...)
$(\cdot,\cdot)$	L^2-scalar product (usually on the domain Ω)
$\|\cdot\|$	L^2-norm
$\|\cdot\|_k$	$H^k(\Omega)$-Sobolev semi norm $\|\cdot\|_k = \|\nabla^k \cdot\|$
$\|\cdot\|_k$	$H^k(\Omega)$-Sobolev norm $\|\cdot\|_k = \sqrt{\sum_{i=0}^{k} \|\cdot\|_i^2}$
$L^2(\Omega)$	The function space of square integrable functions on Ω
$L_0^2(\Omega)$	The same space with average value of zero
$H^s(\Omega)$	Sobolev space of $L^2(\Omega)$-functions with square integrable weak derivative of degree s on Ω
$H_0^1(\Omega)$	Sobolev space of $L^2(\Omega)$-functions with square integrable weak derivative and with trace zero on the boundary
V_h	Usually the finite element space for the velocities
Q_h	Usually the finite element space for the pressure
Re	The Reynolds number
$\mathcal{O}(\cdot), o(\cdot)$	Landau symbols

1 Models

This lecture notes have been assembled for various lectures on computational fluid mechanics, solid mechanics and on fluid-structure interactions. Parts are based on lecture notes on numerical continuum mechanics by Prof. Dr. Rannacher [30]. Most of the material is also found in a book on fluid-structure interactions [31].

1.1 Continuum mechanics

In this chapter, we derive the equations that describe the dynamics of fluids and solids. Matter is composed of molecules, atoms and smaller particles that all interact with each other. A description of the dynamics of these micro-structure is possible by fundamental physical laws. Such a particle centered view-point is however not feasible, if large physical objects are considered that consist of many atoms. To describe every particle in one liter of water, more than 10^{25} molecules must be considered. A description of every single molecule–or even every atom or subatomic particle–in a large scale hydrodynamical problem like the flow of water around a ship is completely out of bounds.

Instead, we consider a *continuum approach* for the description of the large scale dynamics. By a continuum, we denote a volume $V(t) \subset \mathbb{R}^3$ of (different) particles. Instead of describing every single particle, we only observe some few averaged properties of the complete volume. These properties are all considered as local density distributions. As example, we will denote by $\mathbf{v}(x,t)$ the average velocity of whatever particle may be in position $x \in V(t)$ at a given time t. Usually we assume that all physical quantities possess some smoothness. Depending on the situation, we will ask for integrability, continuity or differentiability. Single particles are not distinguished any more. This is called the *continuum hypothesis* or *continuum assumption*.

In the following we will derive fundamental equations that describe the interplay of these averaged quantities. We will distinguish between basic physical principles, the *conservation principles* and *material laws*. While we the conservation principles are based on *first principles* and we think of them as exact, *material laws* are usually simplifications, idealizations and derived by observation and measurements.

1.1.1 Coordinate systems

In the following, by $V(t) \subset \mathbb{R}^3$ we denote a *material volume*. We assume that $V(t)$ is entirely occupied by some material. This material has physical properties like density $\rho : V(t) \to \mathbb{R}$, velocity $\mathbf{v} : V(t) \to \mathbb{R}^3$, which is a three dimensional vector field, temperature $T : V(t) \to \mathbb{R}$ or pressure $p : V(t) \to \mathbb{R}$. We assume that the volume is moving. By $t_0 \in \mathbb{R}$ we denote the *initial time* and we observe the volume for $t \geqslant t_0$. By $V_0 := V(t_0)$ we denote the *reference configuration* of the volume. Often, t_0 is set arbitrarily, but we usually think of a system that is at rest and unstressed, e.g. a container filled with resting fluid or an elastic obstacle that is not deformed and where no stresses act. At time $t \geqslant t_0$, we denote by $V(t)$ the *current configuration*.

The volume $V(t)$ consists of particles, and we call $\hat{V} := V_0$ the material domain. For every particle $\hat{x} \in \hat{V}$, we denote by $x(\hat{x}, t) \in V(t)$ the location of the particle at time $t \geqslant t_0$. We assume that the path $\{x(\hat{x}, t), t \geqslant t_0\} \subset \mathbb{R}^3$ is continuous and that no two different particles $\hat{x}, \hat{x}' \in \hat{V}$ have the same position at any time $t \geqslant t_0$:

$$x(\hat{x}, t) = x(\hat{x}', t) \quad \Leftrightarrow \quad \hat{x} = \hat{x}'.$$

The mapping $\hat{T}(\hat{x}, t) := x(\hat{x}, t)$ is therefore invertible and we define the inverse mapping as $\hat{T}^{-1}(x, t) := \hat{x}(x, t)$. By $\hat{x}(x, t)$ we denote that particle $\hat{x} \in \hat{V}$ that at time $t \geqslant t_0$ takes position $x \in V(t)$.

In a continuum, we assume that no particles are destroyed or created such that the moving volume $V(t)$ is given by all coordinates $x \in \mathbb{R}^3$ that are occupied by a particle $\hat{x} \in \hat{V}$:

$$V(t) = \{x(\hat{x}, t) \in \mathbb{R}^3,\ \hat{x} \in \hat{V}\}.$$

Figure 1.1 shows this fundamental configuration.

We study the motion of volumes and the first fundamental property is the *deformation* of a particle $\hat{x} \in \hat{V}$. We define the *deformation* $\hat{\mathbf{u}}(\hat{x}, t)$ as

$$\hat{\mathbf{u}}(\hat{x}, t) = x(\hat{x}, t) - \hat{x}, \tag{1.1}$$

and its *material velocity* $\hat{\mathbf{v}}(\hat{x}, t)$ as

$$\hat{\mathbf{v}}(\hat{x}, t) := d_t x(\hat{x}, t) = d_t \hat{\mathbf{u}}(\hat{x}, t).$$

This particle system centered viewpoint for describing the dynamics of a continuum $V(t)$ is denoted as *Lagrangian coordinate system* or *Lagrangian framework*. In the Lagrangian system, we observe single particles $\hat{x} \in \hat{V}$ and follow their paths $x(\hat{x}, t) = \hat{x} + \hat{\mathbf{u}}(\hat{x}, t)$ over time. A Lagrangian viewpoint is the natural approach for problems in solid mechanics, where the particles in the reference system are closely linked to each other and where forces are related to the relative deformation of particles to each other (think of a spring). Considering the

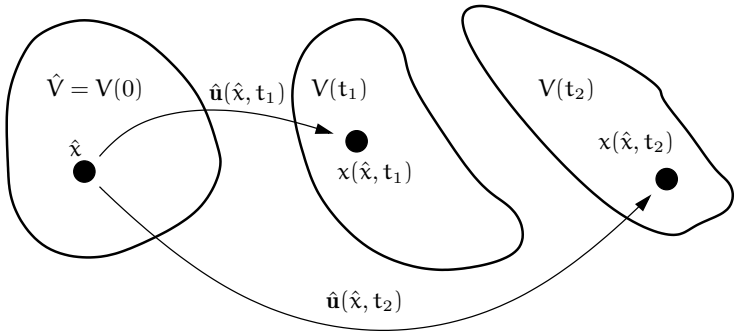


Figure 1.1: The Lagrangian reference system. We describe the path of particles $\hat{x} \in \hat{V}$ over time. The reference volume $\hat{V}$ takes different current configurations $V(t)$ at different times. The particles within $V(t_1)$ are the same particles as in $V(t_2)$ or in $\hat{V} = V(t_0)$.

dynamics of elastic solids a volume comes back to the reference configuration, if the system is free of external forces

$$\begin{array}{c} \hat{V} = V_0 \\ \hat{x} = x(\hat{x}, 0) \end{array} \xrightarrow{\text{external forces act}} \begin{array}{c} V(t) \\ x(\hat{x}, t) \end{array} \xrightarrow{\text{absence of external forces}} \begin{array}{c} V(t_\infty) = \hat{V} \\ \hat{x} = x(\hat{x}, t_\infty) \end{array}$$

Deformation and velocity can also be defined in the current configuration $V(t)$. By

$$x = \hat{x} + \hat{u}(\hat{x}, t) \quad \Leftrightarrow \quad u(x, t) := \hat{u}(\hat{x}, t) = x - \hat{x}$$

we have an expression $u(x, t)$ for the deformation at the spatial location $x \in V(t)$. By $u(x, t)$ we describe the deformation of a particle in location $x \in \mathbb{R}^3$ at time t, we however do not know or determine which individual particle $\hat{x}$ we have in mind. If we describe all quantities in the current configuration $V(t)$ and if we are not interested in single particles at all we do not even need the concept of a reference domain.

The difference between both approaches is the viewpoint: where $\hat{u}(\hat{x}, t)$ denotes the deformation of the particle $\hat{x}$ at time t, by $u(x, t)$ we denote the deformation of whatever particle $\hat{x}$ happens to be at location x at time t. If at time t it holds $x = x(\hat{x}, t)$, both concepts of deformation describe the same configuration. If we base the description of the continuum on the spatial coordinates $x \in V(t)$, we speak of the *Eulerian framework*, where the focus is set on a spatial domain $V \subset \mathbb{R}^3$ and all points $x \in V$, see Figure 1.2. This viewpoint is natural for fluid-dynamical problems. We consider the estimation of the drag-coefficient of a car. Here, the attention is on the flow around the car and we measure forces on the surface of the car, irrespective of the actual particle that at time $t \geqslant 0$ interacts with the car. In fluid dynamics, we want to describe velocity and pressure at spatial points $x \in V$. Usually, we are not interested in what particle interacts with the car and where this particle comes from. Fluids like air or water do not have a *memory*. They behave in the same way regardless of their history. This of course is not true for all liquids. Material like polymers or rubber (which can be

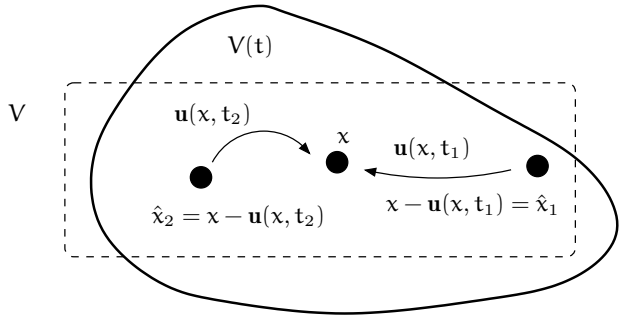


Figure 1.2: The Eulerian reference system. We observe spatial coordinates $x \in V$, where $V \subset \mathbb{R}^3$ is a fixed view. Particles $\hat{x}$ may enter the domain V at a given time and leave it at another time. We observe properties of particles at certain times and locations, we however do not describe and follow the course of individual particles.

described as a fluid, if it is hot) actually do have a memory. Such viscoelastic fluids however are out of the scope of this book.

The Eulerian velocity $\mathbf{v}(x,t)$ is defined as the velocity in position $x \in \mathbb{R}^3$ at time t and given as

$$\mathbf{v}(x,t) = \partial_t \mathbf{u}(x,t) = \partial_t \hat{\mathbf{u}}(\hat{x},t) = \hat{\mathbf{v}}(\hat{x},t).$$

In the Eulerian viewpoint, we do not describe, which particle $\hat{x}$ takes this position.

1.1.2 Deformation gradient

In continuum mechanics, we study the behavior of moving and deforming continua $V(t)$ over time. In the following we describe the relative change of positions $x(\hat{x},t)$ and $x(\hat{y},t)$ of two particles $\hat{x},\hat{y} \in \hat{V}$ in a moving continuum. Relative change of location is called strain, and strain will show to be the most fundamental quantity that causes stress within the material. By stress, we denote the internal forces between the neighboring particles in a continuum.

Let $\hat{x} \in \hat{V}$ and $\hat{y} \in \hat{V}$ be two particles that are infinitesimally close to each other, i.e. $|\hat{y} - \hat{x}| \to 0$. Under deformation, these two particles have the position $x = \hat{x} + \hat{\mathbf{u}}(\hat{x}) \in V$ and $y = \hat{y} + \hat{\mathbf{u}}(\hat{y}) \in V$. We measure the change in position $y - x$ in V with respect to $\hat{y} - \hat{x}$ in $\hat{V}$, see Figure 1.3. By first order Taylor expansion we deduce

$$\begin{aligned} y - x &= \hat{y} + \hat{\mathbf{u}}(\hat{y}) - \hat{x} - \hat{\mathbf{u}}(\hat{x}) \\ &= \hat{y} - \hat{x} + \sum_{i=1}^{d} \hat{\partial}_i \mathbf{u}(\hat{x}) \cdot (\hat{y} - \hat{x}) + O(|\hat{y} - \hat{x}|^2) \\ &= \hat{y} - \hat{x} + \hat{\nabla}\hat{\mathbf{u}}(\hat{x})(\hat{y} - \hat{x}) + O(|\hat{y} - \hat{x}|^2), \end{aligned} \tag{1.2}$$

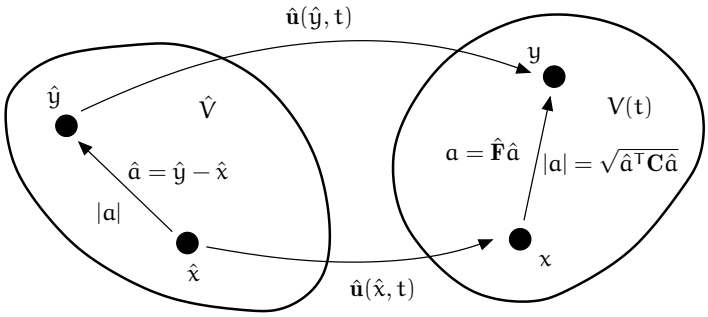


Figure 1.3: Transformation of infinitesimal line segment $\hat{a}$ to a with $|\hat{a}| \to 0$. Deformation gradient $\hat{\mathbf{F}} = I + \hat{\nabla}\hat{\mathbf{u}}$ and squared length change $|a|^2 = \hat{a}^T \hat{\mathbf{C}} \hat{a}$ indicated by the right Cauchy-Green tensor $\hat{\mathbf{C}} = \hat{\mathbf{F}}^T \hat{\mathbf{F}}$.

where by $|\hat{x}| = \sqrt{\sum_{i=1}^{d} \hat{x}_i^2}$ we denote the Euclidean norm, by $\hat{x} \cdot \hat{y} = \sum_{i=1}^{d} \hat{x}_i \hat{y}_i$ the Euclidean scalar product and by $\hat{\partial}_i$ the partial derivative with respect to $\hat{x}_i$ in the Lagrangian coordinate system. Considering the relative change in position, it holds

$$\frac{y - x}{|\hat{y} - \hat{x}|} = [I + \hat{\nabla}\hat{\mathbf{u}}(\hat{x})] \frac{\hat{y} - \hat{x}}{|\hat{y} - \hat{x}|} + O(|\hat{y} - \hat{x}|). \tag{1.3}$$

We define

Definition 1.1 (Deformation Gradient)**.** Let $\hat{\mathbf{u}}$ be a differentiable deformation field in the material volume $\hat{V}$. The *deformation gradient*

$$\hat{\mathbf{F}}(\hat{x}, t) := I + \hat{\nabla}\hat{\mathbf{u}}(\hat{x}, t),$$

denotes the local change of relative position under deformation.

The deformation gradient is the fundamental measure in structure dynamics.

Lemma 1.2 (Determinant of the deformation gradient)**.** Let $\hat{V}$ be a reference volume and $\hat{\mathbf{u}} : \hat{V} \to \mathbb{R}^d$ be a differentiable deformation field. The determinant of the deformation gradient $\hat{J} := \det(\hat{\mathbf{F}})$ denotes the local change of volume:

$$|V(t)| = \int_{\hat{V}} \hat{J} \, dx.$$

Proof. It holds by the transformation theorem

$$|V(t)| = \int_{V(t)} 1 \, dx = \int_{\hat{V}} \det(I + \hat{\nabla}\hat{\mathbf{u}}) \, d\hat{x} = \int_{\hat{V}} \hat{J} \, d\hat{x}.$$

□

The deformation gradient $\hat{\mathbf{F}}$ applies to the Lagrangian viewpoint. For an Eulerian description in $V(t)$, we can define the inverse deformation gradient $\mathbf{F}$ in a similar way. For two spatial coordinates $x, y \in V$ belonging to particles $\hat{x}$ and $\hat{y}$ in $\hat{V}$ it holds

$$\frac{\hat{y}-\hat{x}}{|y-x|} = \mathbf{F}(x)\frac{y-x}{|y-x|} + O(|y-x|),$$

with the *inverse deformation gradient* $\mathbf{F}(x,t) = I - \nabla \mathbf{u}(x,t)$. It holds $\mathbf{F} = \hat{\mathbf{F}}^{-1}$.

Very often, it will be necessary to rapidly switch between different viewpoints on the same physical problem. Sometimes, it is appropriate to consider the material centered reference domain $\hat{V}$, while sometimes the Eulerian viewpoint of the current configuration $V(t)$ is better suited. Usually, we denote all entities in the material system with a hat "$\hat{\cdot}$" and use the same notation without the hat for the Eulerian notation. Every basic property like velocity and deformation has a Eulerian counterpart, e.g. $\mathbf{v}(x,t) = \hat{\mathbf{v}}(\hat{x},t)$ and $\mathbf{u}(x,t) = \hat{\mathbf{u}}(\hat{x},t)$, where for $\hat{x}$ and x at a given time $t \geqslant t_0$ it always holds $x = \hat{x} + \hat{\mathbf{u}}(\hat{x},t)$. When referring to derivatives of these basic quantities, a simple "$\nabla \mathbf{u} = \hat{\nabla}\hat{\mathbf{u}}$" is usually wrong. Instead, we need to derive rules to map between both coordinate frames:

Lemma 1.3 (Transformation between the reference and the current configuration)**.** Let $I = [0,T]$ be a time interval, $\hat{V}$ be a reference domain and $\hat{\mathbf{u}} \in C^1(I \times \hat{V})^3$. We assume that $T := \mathrm{id} + \hat{\mathbf{u}}$ defines a C^1-diffeomorphism between $\hat{V}$ and

$$V(t) = \{\hat{x} + \hat{\mathbf{u}}(\hat{x},t),\ \hat{x} \in \hat{V}\}.$$

Let $\hat{f} \in C^1(I \times \hat{V})$ and $f(x,t) = f(x(\hat{x},t),t) = \hat{f}(\hat{x},t)$ be its counterpart in the current configuration. It holds

$$\hat{\nabla}\hat{f} = \hat{\mathbf{F}}^T \nabla f \tag{1.4}$$

and

$$d_t f = d_t \hat{f}, \quad \partial_t f = \partial_t \hat{f} - \hat{\mathbf{F}}^{-T}\hat{\nabla}\hat{f}\cdot\hat{\mathbf{v}}. \tag{1.5}$$

Let $\hat{\mathbf{w}} \in C^1(I \times \hat{V})^3$ be given with counterpart $\mathbf{w}(x,t) = \hat{\mathbf{w}}(\hat{x},t)$. It holds

$$\hat{\nabla}\hat{\mathbf{w}} = \nabla \mathbf{w}\hat{\mathbf{F}}. \tag{1.6}$$

Proof. For the spatial derivative of $f(x,t)$ it holds with $x(\hat{x},t) = \hat{x} + \hat{\mathbf{u}}(\hat{x},t)$:

$$\hat{\partial}_i \hat{f}(\hat{x},t) = \hat{\partial}_i f(x(\hat{x},t),t) = \sum_j \partial_j f(x,t)\hat{\partial}_i x^j(\hat{x},t) = \sum_j \partial_j f(x,t)\hat{\mathbf{F}}_{ji}.$$

Hence

$$\hat{\nabla}\hat{f} = \hat{\mathbf{F}}^T \nabla f.$$

Then, for a vector field $\mathbf{w} = (\mathbf{w}_i)_i$ it follows

$$(\hat{\nabla}\hat{\mathbf{w}})_{ij} = \hat{\partial}_j \hat{\mathbf{w}}_i = \sum_k \partial_k \mathbf{w}_i \hat{\partial}_j x^k(x,t) = (\nabla \mathbf{w})_{ik}\hat{\mathbf{F}}_{kj} = (\nabla \mathbf{w}\hat{\mathbf{F}})_{ij}$$

For the total time derivatives we get

$$\begin{aligned} d_t f(x,t) &= \partial_t f(x,t) + \nabla f(x,t) \cdot \mathbf{v}(x,t) \\ d_t \hat{f}(\hat{x},t) &= d_t f(\hat{x} + \hat{\mathbf{u}}(\hat{x},t), t) = \partial_t f(x,t) + \nabla f(x,t) \cdot \hat{\mathbf{v}}(\hat{x},t). \end{aligned} \tag{1.7}$$

which, with $\hat{\mathbf{v}}(\hat{x},t) = \mathbf{v}(x,t)$, shows that $d_t f = d_t \hat{f}$. On the other hand it holds

$$d_t \hat{f}(\hat{x},t) = \partial_t \hat{f}(\hat{x},t).$$

This, with (1.7) and (1.4) shows

$$\partial_t f(x,t) = d_t \hat{f}(\hat{x},t) - \nabla f(x,t) \cdot \mathbf{v}(x,t) = \partial_t \hat{f}(\hat{x},t) - \hat{\mathbf{F}}^{-T} \hat{\nabla} \hat{\mathbf{f}}(\hat{x},t) \cdot \hat{\mathbf{v}}(\hat{x},t)$$

which shows the last relation. □

1.1.3 Strain

Strain is defined as the deformation within a body relative to a reference length. Fixed body rotations or translation undergo no strain, as the relative positions of all particles is kept constant. Strain will be the basic quantity used to describe stresses in solid mechanics. A simple model is a spring, where change of length - the strain - will be proportional to a force.

Let $\hat{a} = \hat{y} - \hat{x}$ be the vector of a line-segment between the two points $\hat{x}, \hat{y} \in \hat{V}$. Then, given a deformation field $\hat{\mathbf{u}} : \hat{V} \to \mathbb{R}^3$, let $x = \hat{x} + \hat{\mathbf{u}}(\hat{x})$ and $y = \hat{y} + \hat{\mathbf{u}}(\hat{y})$ and set $a := y - x$. It holds with (1.3) that

$$a = y - x = \hat{\mathbf{F}}(\hat{x})\hat{a} + O(|\hat{a}|^2),$$

and the length of $|a|$ is given as

$$\begin{aligned} |a| &= \sqrt{(\hat{\mathbf{F}}\hat{a}, \hat{\mathbf{F}}\hat{a}) + O(|\hat{a}|^3)} \\ &= |\hat{a}| \sqrt{\left(\hat{\mathbf{F}}\frac{\hat{a}}{|\hat{a}|}, \hat{\mathbf{F}}\frac{\hat{a}}{|\hat{a}|}\right) + O(|\hat{a}|)} \\ &= |\hat{a}| \left(\sqrt{\left(\hat{\mathbf{F}}\frac{\hat{a}}{|\hat{a}|}, \hat{\mathbf{F}}\frac{\hat{a}}{|\hat{a}|}\right)} + O(|\hat{a}|) \right) \\ &= \sqrt{(\hat{a}^T, \hat{\mathbf{F}}^T\hat{\mathbf{F}}\hat{a})} + O(|\hat{a}|^2), \end{aligned}$$

which follows using the Taylor $\sqrt{1+x} = 1 + \mathcal{O}(x)$. For an illustration of the deformation gradient, see Figure 1.3. By $\hat{\mathbf{C}} = \hat{\mathbf{F}}^T\hat{\mathbf{F}}$ we denote the *right Cauchy-Green tensor* which is also denoted as the *Green deformation tensor*. This tensor is symmetric and positive definite, as

$$(\hat{\mathbf{C}}\hat{a}, \hat{a}) = (\hat{\mathbf{F}}\hat{a}, \hat{\mathbf{F}}\hat{a}) = \|\hat{\mathbf{F}}\hat{a}\|^2 > 0 \quad \forall \hat{x} \neq 0,$$

and it describes the (squared) length scaling of a line-segment in direction $\hat{a} = \hat{y} - \hat{x}$. A further commonly used strain measure is the *Green-Lagrange strain tensor* $\hat{\mathbf{E}} := \frac{1}{2}(\hat{\mathbf{C}} - I) =$

$\frac{1}{2}(\hat{\mathbf{F}}^\mathsf{T}\hat{\mathbf{F}} - \mathrm{I})$ that measures the (squared) length change of a line-segment $\hat{a} = \hat{y} - \hat{x}$ under deformation $a = y - x$:

$$\begin{aligned}\frac{1}{2}\left(|a|^2 - |\hat{a}|^2\right) &= \frac{1}{2}\left(\hat{a}^\mathsf{T}\hat{\mathbf{C}}\hat{a} - \hat{a}^\mathsf{T}\hat{a}\right) + O(|\hat{a}|^3) \\ &= \hat{a}^\mathsf{T}\left(\frac{1}{2}(\hat{\mathbf{F}}^\mathsf{T}\hat{\mathbf{F}} - \mathrm{I})\right)\hat{a} + O(|\hat{a}|^3).\end{aligned} \tag{1.8}$$

The tensors $\hat{\mathbf{C}} = \hat{\mathbf{F}}^\mathsf{T}\hat{\mathbf{F}}$ and $\hat{\mathbf{E}} = \frac{1}{2}(\hat{\mathbf{C}} - \mathrm{I})$ are nonlinear functions in the deformation $\hat{\mathbf{u}}$:

$$\hat{\mathbf{C}} = \mathrm{I} + \hat{\nabla}\hat{\mathbf{u}} + \hat{\nabla}\hat{\mathbf{u}}^\mathsf{T} + \hat{\nabla}\hat{\mathbf{u}}^\mathsf{T}\hat{\nabla}\hat{\mathbf{u}}, \quad \hat{\mathbf{E}} = \frac{1}{2}\left(\hat{\nabla}\hat{\mathbf{u}} + \hat{\nabla}\hat{\mathbf{u}}^\mathsf{T} + \hat{\nabla}\hat{\mathbf{u}}^\mathsf{T}\hat{\nabla}\hat{\mathbf{u}}\right).$$

Given a very small variation in deformation, i.e. $|\hat{\nabla}\hat{\mathbf{u}}| \ll 1$, one sometimes uses linearization of the strain tensors as an approximation:

$$\mathbf{c} = \mathrm{I} + \hat{\nabla}\hat{\mathbf{u}} + \hat{\nabla}\hat{\mathbf{u}}^\mathsf{T}, \quad \boldsymbol{\epsilon} = \frac{1}{2}\left(\hat{\nabla}\hat{\mathbf{u}} + \hat{\nabla}\hat{\mathbf{u}}^\mathsf{T}\right).$$

These approximations can be good approximations under certain conditions. One however has to be careful, as having a small deformation $\hat{\mathbf{u}}$ is not a sufficient condition for this linearization.

The tensors $\hat{\mathbf{F}}$, $\hat{\mathbf{C}}$, $\hat{\mathbf{E}}$ and the linearized strain tensor $\boldsymbol{\epsilon}$ all refer to the Lagrangian material coordinate system. They are called *material strain tensors*. Sometimes, we need to express strain in the spatial coordinate system, directly on the current frame $V(t)$. Hence let $x, y \in V(t)$ be two spatial coordinates at time $t \geqslant t_0$, spanning the line-segment $a = y - x$. By $\hat{x}, \hat{y} \in \hat{V}$ we denote the material points corresponding to this line-segment. These span the material line-segment $\hat{a} = \hat{y} - \hat{x}$. Similar to (1.3), but using the Eulerian notation $\mathbf{u}(x, t) = \hat{\mathbf{u}}(\hat{x}, t)$ we get

$$\hat{y} - \hat{x} = y - \mathbf{u}(y) - (x - \mathbf{u}(x)) = [\mathrm{I} - \nabla\mathbf{u}(x)](y - x) + O(|y - x|^2).$$

By $\mathbf{F}(x) = \mathrm{I} - \nabla\mathbf{u}(x)$ we denote the *inverse deformation tensor*. It holds $\mathbf{F}(x) = \hat{\mathbf{F}}(\hat{x})^{-1}$ for $x = \hat{x} + \hat{\mathbf{u}}(\hat{x})$. $\mathbf{F}(x)$ is the deformation gradient in the current configuration and it acts on the spatial coordinate system. With help of $\mathbf{F} = \mathrm{I} - \nabla\mathbf{u}$ we can immediately analyze length changes in the spatial system. Let $a = y - x$ and $\hat{a} = \hat{y} - \hat{x}$. It holds

$$|\hat{a}|^2 = (\mathbf{F}a, \mathbf{F}a) + O(|a|^3) = a^\mathsf{T}\mathbf{F}^\mathsf{T}\mathbf{F}a + O(|a|^3) = a^\mathsf{T}\hat{\mathbf{F}}^{-\mathsf{T}}\hat{\mathbf{F}}^{-1}a + O(|a|^3).$$

The tensor $\mathbf{b}^{-1} := \hat{\mathbf{F}}^{-\mathsf{T}}\hat{\mathbf{F}}^{-1} = \mathbf{F}^\mathsf{T}\mathbf{F}$ is the inverse of the *left Cauchy-Green tensor* $\mathbf{b}$

$$\mathbf{b} = \hat{\mathbf{F}}\hat{\mathbf{F}}^\mathsf{T}.$$

As $\hat{\mathbf{C}}$, $\mathbf{b}$ is symmetric positive definite. Finally, we can define the spatial Eulerian counterpart $\mathbf{e} = \frac{1}{2}(\mathrm{I} - \mathbf{F}^\mathsf{T}\mathbf{F})$ to the Cauchy-Green strain tensor $\hat{\mathbf{E}}$. By (1.8), it holds (we neglect higher order terms)

$$\frac{1}{2}\left(|a|^2 - |\hat{a}|^2\right) \approx \hat{a}^\mathsf{T}\left(\frac{1}{2}\left(\hat{\mathbf{F}}^\mathsf{T}\hat{\mathbf{F}} - \mathrm{I}\right)\right)\hat{a}$$

which transform by using $\hat{a} \approx \hat{\mathbf{F}}^{-1} a$ to a Eulerian description

$$\frac{1}{2}\left(|a|^2 - |\hat{a}|^2\right) \approx a^{\mathsf{T}} \hat{\mathbf{F}}^{-\mathsf{T}} \left(\frac{1}{2}\left(\hat{\mathbf{F}}^{\mathsf{T}}\hat{\mathbf{F}} - \mathrm{I}\right)\right) \hat{\mathbf{F}} a = a^{\mathsf{T}} \left(\frac{1}{2}\left(\mathrm{I} - \hat{\mathbf{F}}^{-\mathsf{T}}\hat{\mathbf{F}}\right)\right) a = a^{\mathsf{T}} \left(\frac{1}{2}\left(\mathrm{I} - \mathbf{F}^{\mathsf{T}}\mathbf{F}\right)\right) a$$

We introduce

$$\mathbf{e} := \frac{1}{2}\left(\mathrm{I} - \mathbf{F}^{\mathsf{T}}\mathbf{F}\right) = \frac{1}{2}\left(\mathrm{I} - \hat{\mathbf{F}}^{-\mathsf{T}}\hat{\mathbf{F}}^{-1}\right) = \hat{\mathbf{F}}^{-\mathsf{T}}\hat{\mathbf{E}}\hat{\mathbf{F}}^{-1}$$

the symmetric *Euler-Almansi strain tensor* $\mathbf{e}$ that enables us to relate length changes to the Eulerian line segment a:

$$\frac{1}{2}\left(|a|^2 - |\hat{a}|^2\right) = a^{\mathsf{T}} \mathbf{e} a + \mathrm{O}(|a|^3).$$

If for a body $\hat{V}$ it holds $\hat{\mathbf{C}} = \mathrm{I}$ it follows that $\hat{\mathbf{E}} = 0$, and no relative changes in the position of material points $\hat{x}$ and $\hat{y}$ occur. Lengths and angles are maintained. A material body that can only undergo motion with $\hat{\mathbf{E}} = 0$ is called a *rigid body*.

Remark 1.4 (Right Cauchy-Green or Green-Lagrange strain tensor)**.** We have two different strain measures at hand. The right Cauchy-Green strain tensor $\hat{\mathbf{C}}$ and the Green-Lagrange strain tensor $\hat{\mathbf{E}}$. Both are firmly linked and can be used to describe strains caused by deformation. For describing material laws, we will derive models, that characterize the materials reaction on strain. Most simple models will assume a linear dependency between strain and stress: if no strain is given, no stress is induced. Here, the Green-Lagrange strain tensor $\hat{\mathbf{E}}$ is the better basis, as $\hat{\mathbf{E}} = 0$ denotes a no-strain condition and a linear function $f(\hat{\mathbf{E}})$ can be consulted to model the strain-stress relationship. △

1.1.4 Rate of deformation and strain rate

The strain tensor is a fundamental quantity in solid mechanics, where we assume that a finite force will cause a finite deformation. An ideal spring will linearly react on external forces by some finite extension, which directly refers to strain. In fluid-mechanics however finite forces can lead to infinite deformation. A river, which is driven by the constant gravity force causes infinite strain, although the force is bounded. Here it is not the deformation and the deformation gradient that is of interest; but it is its temporal variation that serves as key quantity to model the internal forces (stresses) of the material. We already discussed that for fluid-dynamical observations, the Eulerian viewpoint is more meaningful. Hence we will derive a measure for the rate of strain in the current system $V(t)$.

By $\hat{x}, \hat{y} \in \hat{V}$ we denote two material points spanning the line-segment $\hat{a} = \hat{y} - \hat{x}$. We follow their positions $x(t) = \hat{x} + \hat{\mathbf{u}}(\hat{x}, t) \in V(t)$, $y(t) = \hat{y} + \hat{\mathbf{u}}(\hat{y}, t) \in V(t)$ and the resulting line-segment $a(t) = y(t) - x(t)$ in the current configuration $V(t)$. With $a(t) = \hat{\mathbf{F}}(t)\hat{a}$ it holds

$$\partial_t a(t) = \partial_t \hat{\mathbf{F}}(t)\hat{a}, \tag{1.9}$$

and for the deformation gradient $\hat{\mathbf{F}}(t) = \mathrm{I} + \hat{\nabla}\hat{\mathbf{u}}(t)$ we get

$$\partial_t \hat{\mathbf{F}} = \partial_t \hat{\nabla}\hat{\mathbf{u}} = \hat{\nabla}\hat{\mathbf{v}}.$$

where we assumed sufficient regularity to change the order of derivatives. By $\hat{\nabla}\hat{\mathbf{v}}$ we denote the *material velocity gradient*. The material velocity gradient $\hat{\nabla}\hat{\mathbf{v}}(\hat{x}, t)$ denotes the spatial change of the velocity as given in the Lagrangian material system. The *spatial velocity gradient* $\nabla\mathbf{v}(x, t)$ refers to the spatial change of the velocity of whatever particles are at location x at time t. For $\hat{\mathbf{v}}(\hat{x}) = \mathbf{v}(x)$ with $x = x(\hat{x}) = \hat{x} + \hat{\mathbf{u}}(\hat{x})$ it further holds

$$\partial_t \hat{\mathbf{F}} = \nabla\mathbf{v}\hat{\nabla}x = \nabla\mathbf{v}\hat{\mathbf{F}}.$$

Then, to continue with (1.9)

$$\partial_t a(t) = \nabla\mathbf{v}\hat{\mathbf{F}}\hat{a} = \nabla\mathbf{v}a(t),$$

and the rate of length change is given by

$$\partial_t |a(t)|^2 = (\nabla\mathbf{v}a(t), a(t)) + (a(t), \nabla\mathbf{v}a(t)) = 2\left(\frac{1}{2}(\nabla\mathbf{v} + \nabla\mathbf{v}^T)a(t), a(t)\right).$$

Definition 1.5 (Strain rate tensor)**.** By

$$\dot{\boldsymbol{\epsilon}}(x, t) = \frac{1}{2}\left\{\nabla\mathbf{v}(x, t) + \nabla\mathbf{v}(x, t)^T\right\}.$$

we denote the *strain rate* tensor or the *rate of strain* tensor. It denotes the local change of velocity in the current system.

1.1.5 Stress

Deformation, strain and strain rate are kinematic properties. They simply describe the relative motion of particles within a volume. As such, they are pure observations of the situations and do not depend on the model under consideration. We assume that a material will react on strain or the strain rate. For expanding a spring, a certain force will be necessary.

By *stress* we denote the internal force that is acting on an imaginary surface within the volume $V(t)$. The unit of stress is force per area.

In Figure 1.4 we show a volume $V(t)$ that is cut at an inner surface $S \subset V(t)$. By $x \in S \subset V(t)$ we denote a point on this surface with normal $\mathbf{n}$. The average forces acting on a neighborhood of $x \in S$ is denoted by the *Cauchy traction vector* $\mathbf{t}$. The right sketch of the figure shows this setting in the reference system, where by $\hat{x} \in \hat{S} \subset \hat{V}$ we denote point, surface and volume in reference state. Here, the normal vector is indicated by $\hat{\mathbf{n}}$ and the resulting *first Piola-Kirchhoff traction vector* $\hat{\mathbf{t}}$:

$$\mathbf{t} = \mathbf{t}(x, t, \mathbf{n}), \quad \hat{\mathbf{t}} = \hat{\mathbf{t}}(\hat{x}, t, \hat{\mathbf{n}}).$$

By ds we denote an infinitesimal neighborhood of x on the surface $S \subset V(t)$ and by $d\hat{s}$ the corresponding infinitesimal neighborhood of $\hat{x}$ on $\hat{S} \subset \hat{V}$. Then, it holds

$$\mathbf{t}ds = \hat{\mathbf{t}}d\hat{s},$$

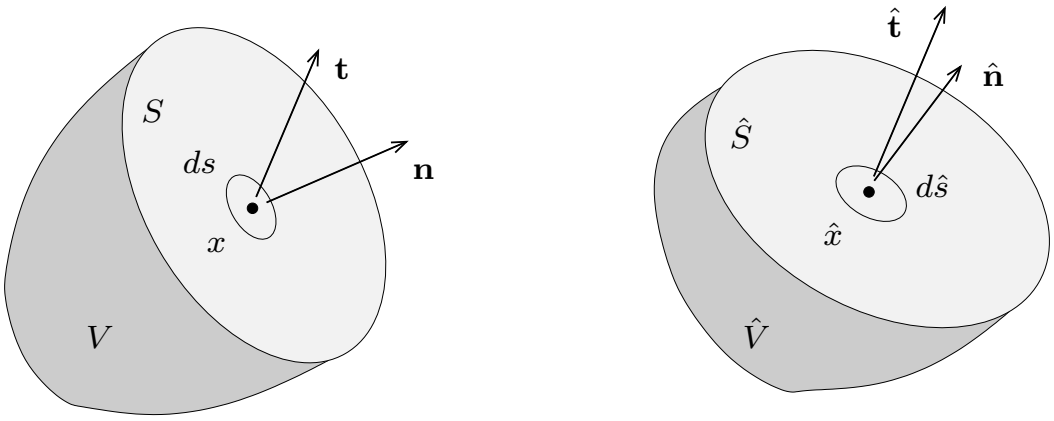


Figure 1.4: Traction vectors on a imaginary surface in the current system (left) and the reference system (right). Cauchy's stress theorem postulates a linear dependency of the traction vectors on the normals $\mathbf{t} = \boldsymbol{\sigma}\mathbf{n}$ and $\hat{\mathbf{t}} = \hat{\mathbf{P}}\hat{\mathbf{n}}$.

such that both traction vectors refer to forces in the current configuration $V(t)$. While $\mathbf{t}$ is a function in variables x and $\mathbf{n}$ of the current configuration, the first Piola-Kirchhoff traction vector is a function of $\hat{x}$ and $\hat{\mathbf{n}}$ in the Lagrangian reference system. Usually, it does not hold $|\mathbf{t}| = |\hat{\mathbf{t}}|$. The unit of stress is *force by area* and $\mathbf{t}$ refers to the area of a domain surface ds while $\hat{\mathbf{t}}$ refers to the area of the undeformed reference surface $d\hat{s}$.

The traction vectors describe a *surface tension*. Such surface tensions arise from friction or contact. Another example for a surface tension is the pressure in a liquid or gas that pushes the particle to each other (or apart from each other).

The surface tensions depend on the normal vector $\mathbf{n}$ of the imaginary surface. It holds

Theorem 1.6 (Cauchy's stress theorem)**.** There exist unique second order tensors $\boldsymbol{\sigma}$ and $\hat{\mathbf{P}}$, such that

$$\mathbf{t}(x, t, \mathbf{n}) = \boldsymbol{\sigma}(x, t)\mathbf{n}, \quad \hat{\mathbf{t}}(\hat{x}, t, \hat{\mathbf{n}}) = \hat{\mathbf{P}}(\hat{x}, t)\hat{\mathbf{n}}.$$

The tensor $\boldsymbol{\sigma} = \boldsymbol{\sigma}^T$ is symmetric and called the *Cauchy stress tensor*, the tensor field $\hat{\mathbf{P}}$ is called the *first Piola-Kirchhoff stress tensor*. $\hat{\mathbf{P}}$ is usually not symmetric.

Proof. For the proof, we refer to the literature [12]. □

One immediate consequence of Cauchy's stress theorem is that traction vectors for opposite normal vectors annihilate each other, Newton's law of *actio = reactio*

$$\mathbf{t}(x, t, -\mathbf{n}) = \boldsymbol{\sigma}(x, t)(-\mathbf{n}) = -\boldsymbol{\sigma}(x, t)\mathbf{n} = -\mathbf{t}(x, t, \mathbf{n}).$$

As the Cauchy stress tensor must be symmetric, it consists of six independent components

$$\boldsymbol{\sigma} = \begin{pmatrix} \boldsymbol{\sigma}_{11} & \boldsymbol{\sigma}_{12} & \boldsymbol{\sigma}_{13} \\ \boldsymbol{\sigma}_{12} & \boldsymbol{\sigma}_{22} & \boldsymbol{\sigma}_{23} \\ \boldsymbol{\sigma}_{13} & \boldsymbol{\sigma}_{23} & \boldsymbol{\sigma}_{33} \end{pmatrix}.$$

The second order tensor $\hat{\mathbf{P}}$ is usually not symmetric and consists of nine independent entries. For the relation of $\boldsymbol{\sigma}$ and $\hat{\mathbf{P}}$ it holds

$$\boldsymbol{\sigma}\mathbf{n}ds = \hat{\mathbf{P}}\hat{\mathbf{n}}d\hat{s},$$

such that the two different traction vectors describe the transformation of a surface integral. We will get back to this relation in Section 1.1.7.

The components of the stress tensor are best understood by a decomposition of stresses into *normal stress* $\sigma \in \mathbb{R}$ and *shear stress* $\tau \in \mathbb{R}$. Let $\mathbf{t}$ be a stress vector in $x \in V(t)$ on a imaginary surface S with normal vector $\mathbf{n}$. The normal-stress σ is defined as the projection of the traction vector in normal direction

$$\sigma = \mathbf{t}^T\mathbf{n} = (\mathbf{n}, \boldsymbol{\sigma}\mathbf{n}),$$

while the shear stress is defined as the tangential part of the stress

$$\tau = \mathbf{t}^T\mathbf{t}_1 = (\mathbf{t}_1, \boldsymbol{\sigma}\mathbf{n}),$$

where $\mathbf{t}_1$ is the tangential vector that arises from projection of $\mathbf{t}$ onto the surface

$$\mathbf{t}_1 = \frac{\mathbf{t} - \boldsymbol{\sigma}\mathbf{n}}{\|\mathbf{t} - \boldsymbol{\sigma}\mathbf{n}\|}.$$

Then, the stress vector can be decomposed into the normal stress σ and shear stress τ by

$$\mathbf{t} = (\mathbf{t}, \mathbf{n})\mathbf{n} + (\mathbf{t}, \mathbf{t}_1)\mathbf{t}_1 = \sigma\mathbf{n} + \tau\mathbf{t}_1.$$

Here $\sigma, \tau \in \mathbb{R}$ are the lengths of the stress vectors in normal direction and tangential direction. Given the Cauchy stress tensor $\boldsymbol{\sigma}$, it holds

$$\sigma = (\mathbf{n}, \boldsymbol{\sigma}\mathbf{n}), \quad \tau = (\mathbf{t}, \boldsymbol{\sigma}\mathbf{n}).$$

If for the normal stress it holds $\sigma < 0$, the material undergoes a compression, while for $\sigma > 0$ an expansion is given. Further, it holds

$$|\boldsymbol{\sigma}\mathbf{n}|^2 = |\mathbf{t}|^2 = |\mathbf{n} \cdot \boldsymbol{\sigma}|^2 = \tau^2 + \sigma^2.$$

Next, let us assume that the imaginary surface has normal vector $\mathbf{n} = \mathbf{e}_i$ with $(\mathbf{e}_i)_j = \delta_{ij}$. The normal stress is given

$$\sigma = (\mathbf{e}_i, \boldsymbol{\sigma}\mathbf{e}_i) = \boldsymbol{\sigma}_{ii},$$

by the diagonal entry of the Cauchy-stress tensor, while the shear stress in $\mathbf{e}_k$ direction for $k \neq i$ gets

$$\tau = (\mathbf{e}_k, \boldsymbol{\sigma}\mathbf{e}_i) = \boldsymbol{\sigma}_{ki} = \boldsymbol{\sigma}_{ki}.$$

Hence the diagonal entries of $\boldsymbol{\sigma}$ refer to the normal stresses, while all off-diagonals refer to tangential shear stresses.

Remark 1.7 (Stress in the reference system)**.** Usually only static stresses act in the initial reference state of a system at reference time t_0. In case of a resting fluid, this stress can be caused by the hydrostatic pressure. Sometimes however, initial configurations cannot be considered to be stress-free. An example could be organic material like wood, where the undeformed reference system may be subject to stress caused by growth, see [24]. △

1.1.6 Conservation principles

The most important physical conservation principles in the context of fluid-mechanics and structure-mechanics are conservation of mass, which says that

> mass is neither created nor destroyed,

conservation of momentum that says that

> the change in momentum is equivalent to the force acting

and conservation of angular momentum, saying that

> the change in angular momentum is equal to the torque.

Using the notation derived in the previous section, conservation of mass reads

$$d_t m(V(t)) = 0, \tag{1.10}$$

where the volume's mass $m(V(t))$ is given by

$$m(V(t)) = \int_{V(t)} \rho(x,t)\,dx,$$

with a density ρ. Conservation of momentum gets

$$d_t I(V(t)) = K(V(t)) + K(\partial V(t)), \tag{1.11}$$

with the momentum $I(V(t))$

$$I(V(t)) = \int_{V(t)} \rho(x,t)\mathbf{v}(x,t),dx,$$

and volume and surface forces $K(V(t))$ and $K(\partial V(t))$ given by:

$$K(V(t)) = \int_{V(t)} \rho(x,t)\mathbf{f}(x,t)\,dx, \quad K(\partial V(t)) = \int_{\partial V(t)} \mathbf{t}\,ds.$$

Here, $\mathbf{f}$ is a prescribed volume force density and $\mathbf{t}$ denotes the surface stress in direction $\mathbf{n}$. As discussed, it holds by Cauchy's Stress Theorem 1.6 that this surface force linearly depends on the normal direction such that it can be expressed with help of a stress tensor

$\boldsymbol{\sigma} \in \mathbb{R}^{n\times n}$ as $\mathbf{t} = \boldsymbol{\sigma}\mathbf{n}$. The surface allows for a transformation to a volume integral via the divergence theorem

$$K(\partial V(t)) = \int_{\partial V(t)} \mathbf{n} \cdot \boldsymbol{\sigma}\, ds = \int_{V(t)} \operatorname{div}(\boldsymbol{\sigma})\, dx.$$

Finally, conservation of angular momentum is given by

$$d_t L(V(t)) = D(V(t)), \tag{1.12}$$

where the angular momentum $L(V(t))$ with respect to the origin is given as

$$L(V(t)) = \int_{V(t)} x \times (\rho \mathbf{v})\, dx,$$

and the torque $D(V(t))$ is defined by

$$D(V(t)) = \int_{V(t)} x \times (\rho \mathbf{f})\, dx + \int_{\partial V(t)} x \times (\mathbf{n} \cdot \boldsymbol{\sigma})\, ds.$$

Since the integration domain $V(t)$ in (1.10), (1.11) and (1.12) depends on time t, evaluation of derivatives like $d_t m(V(t))$ is not straightforward and will be accomplished with help of the essential *Reynolds' Transport Theorem*

Theorem 1.8 (Reynolds' Transport Theorem)**.** Let $V(t) \subset \mathbb{R}^d$ be a material volume. Further, let $\Phi(x,t)$ be a differentiable scalar function defined on $V(t)$. Then, it holds

$$d_t \int_{V(t)} \Phi(x,t)\, dx = \int_{V(t)} \left(\partial_t \Phi(x,t) + \operatorname{div}(\Phi \mathbf{v})\right) dx.$$

Before giving a proof to this theorem, we need the following results.

Lemma 1.9 (Partial derivatives of inverse and determinant)**.** Let $\hat{\mathbf{F}} : \mathbb{R} \times \mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n}$ be a differentiable matrix function with differentiable inverse. Let $\hat{J} = \det(\hat{\mathbf{F}})$ be its determinant and $\hat{\mathbf{F}} = (\hat{\mathbf{F}}_{ij})_{ij}$ its elements. It holds

$$\begin{aligned} &(i) && [\hat{\mathbf{F}}^{-1}]' = -\hat{\mathbf{F}}^{-1}\hat{\mathbf{F}}'\hat{\mathbf{F}}^{-1} \\ &(ii) && \frac{\partial J(\hat{x},t)}{\partial \mathbf{F}_{ij}} = \Delta_{ij}, \end{aligned} \tag{1.13}$$

where by

$$\Delta_{ij} = (-1)^{i+j} \det(\hat{\mathbf{F}}_{kl})_{k\neq i, l\neq j}$$

we denote the *cofactor* of $\hat{\mathbf{F}}$.

Proof. (*i*) This first relation follows by observing

$$\hat{\mathbf{F}}\hat{\mathbf{F}}^{-1} = \mathrm{I} \quad \Rightarrow \quad (\hat{\mathbf{F}}\hat{\mathbf{F}}^{-1})' = \hat{\mathbf{F}}[\hat{\mathbf{F}}^{-1}]' + \hat{\mathbf{F}}'\hat{\mathbf{F}}^{-1} \quad \Rightarrow \quad [\hat{\mathbf{F}}^{-1}]' = -\hat{\mathbf{F}}^{-1}\hat{\mathbf{F}}'\hat{\mathbf{F}}^{-1}.$$

(*ii*) For the cofactor

$$\Delta_{ij} := (-1)^{i+j} \det(\hat{\mathbf{F}}_{kl})_{k\neq i, l\neq j}$$

it holds

$$\delta_{ik}\hat{J} = \sum_{l=1}^{n} \Delta_{il}\hat{\mathbf{F}}_{kl}, \quad i = 1, \dots, n.$$

Then, differentiation w.r.t. $\hat{\mathbf{F}}_{ij}$ gives

$$\frac{\partial \hat{J}}{\partial \hat{\mathbf{F}}_{ij}} = \sum_{l=1}^{n} \underbrace{\frac{\partial \Delta_{il}}{\partial \hat{\mathbf{F}}_{ij}}\hat{\mathbf{F}}_{il}}_{=0} + \Delta_{il}\underbrace{\frac{\partial \hat{\mathbf{F}}_{il}}{\partial \hat{\mathbf{F}}_{ij}}}_{=\delta_{lj}} = \Delta_{ij}$$

□

Proof of Theorem 1.8. The map

$$\hat{T} : \hat{x} \mapsto x(\hat{x}, t)$$

is uniquely defined and differentiable with $\hat{\mathbf{F}} := \hat{\nabla}\hat{T}$ and

$$\hat{J}(\hat{x}, t) := \det(\hat{\mathbf{F}}(\hat{x}, t)) > 0$$

By mapping of $V(t)$ to $\hat{V} = V(0)$ it holds

$$\int_{V(t)} \Phi(x, t)dx = \int_{\hat{V}} \Phi(\hat{T}(\hat{x}, t), t)\hat{J}(\hat{x}, t)d\hat{x}.$$

The domain $\hat{V}$ does not depend on the time t. Therefore we can exchange differentiation (with respect to t) and integration (with respect to x) to get

$$\frac{d}{dt}\int_{V(t)} \Phi(x, t)dx = \int_{\hat{V}} \Big\{ d_t\Phi(\hat{T}(\hat{x}, t), t)\hat{J}(\hat{x}, t) + \Phi(\hat{T}(\hat{x}, t), t)\partial_t\hat{J}(\hat{x}, t) \Big\} dx.$$

Then, we get for the (most complicated) derivative of the determinant

$$\partial_t\hat{J}(\hat{x}, t) = \sum_{ij} \frac{\partial \hat{J}}{\partial \hat{\mathbf{F}}_{ij}}\frac{\partial \hat{\mathbf{F}}_{ij}}{\partial t} = \sum_{ij} \frac{\partial \hat{J}}{\partial \hat{\mathbf{F}}_{ij}}\frac{\partial \hat{v}_i}{\partial \hat{x}_j} = \sum_{ijk} \frac{\partial \hat{J}}{\partial \hat{\mathbf{F}}_{ij}}\frac{\partial \mathbf{v}_i}{\partial x_k}\underbrace{\frac{\partial x_k}{\partial \hat{x}_j}}_{=:\hat{\mathbf{F}}_{kj}}$$

and using (1.13) from the proof to Lemma 1.9

$$\partial_t\hat{J}(\hat{x}, t) = \sum_{ik} \underbrace{\Big(\sum_j \Delta_{ij}\hat{\mathbf{F}}_{kj} \Big)}_{=:\delta_{ik}\hat{J}} \frac{\partial v_i}{\partial x_k} = \sum_i \hat{J}\operatorname{div} \mathbf{v}.$$

The further terms get

$$d_t\Phi(\hat{T}(\hat{x},t),t) = \partial_t\Phi(x,t) + \partial_t\hat{T}(x,t)\cdot\nabla\Phi(x,t) = \partial_t\Phi + \mathbf{v}\cdot\nabla\Phi$$

and altogether we have

$$\frac{d}{dt}\int_{V(t)}\Phi(x,t)dx = \int_{\hat{V}}\hat{J}\Big(\partial_t\Phi + \mathbf{v}\cdot\nabla\Phi + \Phi\operatorname{div}\mathbf{v}\Big)d\hat{x}$$

Using

$$\operatorname{div}(\mathbf{v}\Phi) = \Phi\operatorname{div}\mathbf{v} + \mathbf{v}\cdot\nabla\Phi$$

and mapping back to $V(t)$ gives the result. □

Applying this theorem to the scalar value $\Phi(x,t) := \rho(x,t)$ we derive the *Law of Mass Conservation*:

$$\int_{V(t)}\partial_t\rho + \operatorname{div}(\rho\mathbf{v})\,dx = 0.$$

This equation is valid for every volume $V(t)$. Assuming that the expression $\partial_t\rho + \operatorname{div}(\rho\mathbf{v})$ is continuous (which is an assumption on the physical properties of the material), the equation of mass-conservation holds in a point-wise manner

$$\partial_t\rho + \operatorname{div}(\rho\mathbf{v}) = 0. \tag{1.14}$$

The second basic rule is *conservation of momentum*, derived by the scalar values $\Phi(x,t) := \rho(x,t)\mathbf{v}_i(x,t)$ for every component of the velocity field. With a column-wise representation of the stress-tensor $\boldsymbol{\sigma} = (\boldsymbol{\sigma}_1,\dots,\boldsymbol{\sigma}_d)$ Reynolds transport theorem yields:

$$\int_{V(t)}\partial_t(\rho\mathbf{v}_i) + \operatorname{div}(\rho\mathbf{v}_i\mathbf{v})\,dx = \int_{V(t)}\rho\mathbf{f}_i + \operatorname{div}(\boldsymbol{\sigma}_i)\,dx, \quad i = 1,\dots,d.$$

Given continuity of the integrand we can again deduce a point-wise equation

$$\partial_t(\rho\mathbf{v}_i) + \operatorname{div}(\rho\mathbf{v}_i\mathbf{v}) = \rho\mathbf{f}_i + \operatorname{div}(\boldsymbol{\sigma}_i), \quad i = 1,\dots,d.$$

By introducing the external product of two vectors

$$\mathbf{v}\otimes\mathbf{w} \in \mathbb{R}^{d\times d}, \quad (\mathbf{v}\otimes\mathbf{w})_{ij} := \mathbf{v}_i\mathbf{w}_j,$$

we can formulate the equation for the conservation of momentum in *conservative formulation*

$$\partial_t(\rho\mathbf{v}) + \operatorname{div}(\rho\mathbf{v}\otimes\mathbf{v}) = \rho\mathbf{f} + \operatorname{div}(\boldsymbol{\sigma}).$$

Combining this equation with the mass-conservation, we can further deduce the equation for conservation of momentum in the *non-conservative formulation*

$$\rho\partial_t\mathbf{v} + \rho(\mathbf{v}\cdot\nabla)\mathbf{v} = \rho\mathbf{f} + \operatorname{div}(\boldsymbol{\sigma}). \tag{1.15}$$

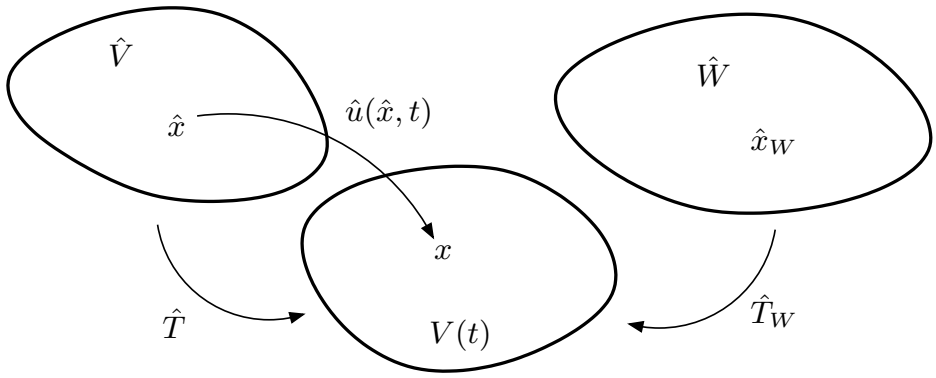


Figure 1.5: Moving Eulerian volume $V(t)$ with Lagrangian reference $\hat{V}$ and third arbitrary reference volume $\hat{W}$.

The equation for the *conservation of angular momentum* is given by

$$d_t \int_{V(t)} x \times (\rho \mathbf{v})\, dx = \int_{V(t)} x \times (\rho \mathbf{f})\, dx + \int_{\partial V(t)} x \times (\mathbf{n} \cdot \boldsymbol{\sigma})\, ds.$$

Applying Reynolds transport theorem we can deduce the following three equations

$$\begin{aligned} i = 1 \quad & \boldsymbol{\sigma}_{23} - \boldsymbol{\sigma}_{32} = 0 \\ i = 2 \quad & \boldsymbol{\sigma}_{13} - \boldsymbol{\sigma}_{31} = 0 \\ i = 3 \quad & \boldsymbol{\sigma}_{12} - \boldsymbol{\sigma}_{21} = 0, \end{aligned}$$

that impose the symmetry of the Cauchy stress tensor

$$\boldsymbol{\sigma} = \boldsymbol{\sigma}^T. \tag{1.16}$$

Further conservation principles are important if physical properties like entropy, energy and temperature are taken into consideration. Since we will deal with isentropic materials only, where all dynamical processes will take place without change of entropy, the three fundamental principles of mass-, momentum- and angular momentum-conservation will be sufficient to describe all desired behavior.

It remains to describe the tensor of surface-forces $\boldsymbol{\sigma}$. This tensor will heavily depend on the material under consideration, whether it is a fluid or a solid, whether the fluid is water, air or blood, the solid may be elastic or plastic or have properties of both. Here, physical modeling comes into place, exact laws for the dependence of this tensor on quantities like velocity and density usually do not exist. Since we know that $\boldsymbol{\sigma}$ is symmetric, six additional equations are required for its description. Stress models will be discussed in Section 1.2.

1.1.7 Conservation principles in different coordinate systems

In this section, we discuss the transformation of the conservation equations, which have been derived in the Eulerian framework, to different coordinate frameworks. Introducing the basic concepts for solid mechanics we already argued that a Lagrangian viewpoint is more natural.

Let $V(t)$ be the moving Eulerian framework and let $\hat{V}$ be the Lagrangian reference system. Further, by $\hat{W}$ we denote an arbitrary second fixed reference system, see Figure 1.5. While the case $\hat{W} = \hat{V}$ is possible, we will allow for arbitrary systems without physical meaning. However, we assume that $\hat{W}$ is fixed in time and that there exists an invertible mapping $\hat{T}_W(t) : \hat{W} \to V(t)$ with gradient $\hat{\mathbf{F}}_W := \hat{\nabla}\hat{T}_W$ and determinant $\hat{J}_W := \det(\hat{\mathbf{F}}_W) > 0$. If we talk about the gradient $\hat{\mathbf{F}}_W$ we request that the mapping $\hat{T}_W$ is differentiable with respect to the spatial variables. We further assume that $\hat{T}_W$ is differentiable with respect to the temporal variable and that the inverse of the mapping $\hat{T}_W^{-1}$ is also differentiable. In other words, $\hat{T}_W$ is assumed to be a C^1-diffeomorphism on $I \times \hat{W}$.

By introducing an arbitrary reference systems $\hat{W}$ we have to deal with three different systems: the Lagrangian particles, $\hat{x} \in \hat{V}$, their Eulerian path $x(\hat{x}, t) \in V(t)$ and further the arbitrary framework with $\hat{x}_W \in \hat{W}$ with $\hat{T}_W(\hat{x}_W, t) = x = \hat{T}(\hat{x}, t)$. Note that it does not hold $\partial_t \hat{T}_W = \hat{\mathbf{v}}$, as we have to distinguish between the physical velocity $\hat{\mathbf{v}}$ of the particles and the velocity $\partial_t \hat{T}_w$ of the arbitrary coordinate system motion.

We start by describing basic properties used to map between the two systems $\hat{W}$ and $V(t)$. First, we introduce the inverse mapping $T_W(t) : V(t) \mapsto \hat{W}$.

Lemma 1.10 (Inverse mapping)**.** By $T_W(t) : V(t) \to \hat{W}$ we denote the *inverse mapping*, by $\mathbf{F}_W := \nabla T_W$ its gradient and by $J_W := \det(\mathbf{F}_W)$ its determinant. Given sufficient regularity, It holds

$$\mathbf{F}_W = \hat{\mathbf{F}}_W^{-1}, \quad J_W := \hat{J}_W^{-1}, \quad \partial_t T_W = -\hat{\mathbf{F}}_W^{-1}\partial_t \hat{T}_W.$$

Proof. It holds

$$T_W \circ \hat{T}_W = \widehat{\mathrm{id}} \quad \Rightarrow \quad \mathbf{F}_W \hat{\mathbf{F}}_W = I \quad \Rightarrow \quad \mathbf{F}_W = \hat{\mathbf{F}}_W^{-1}.$$

By taking the determinant of both sides, we immediately get $J_W = \hat{J}_W^{-1}$. Finally,

$$T_W \circ \hat{T}_W = \widehat{\mathrm{id}} \quad \Rightarrow \quad 0 = d_t T_W(\hat{T}_W(\hat{x}, t), t) = \partial_t T_W + \nabla T_W \partial_t \hat{T}_W.$$

Using $\nabla T_W = \mathbf{F}_W = \hat{\mathbf{F}}_W^{-1}$ we obtain the relation $\partial_t T_W = -\hat{\mathbf{F}}_W^{-1}\partial_t \hat{T}_W$. □

In Lemma 1.3 we already considered the transformation of spatial and temporal derivatives between the Eulerian and the Lagrangian coordinate system. Similarly it holds for a scalar function $f : V(t) \to \mathbb{R}$ and a vector field $\mathbf{w} : V(t) \to \mathbb{R}^d$ with counterparts $\hat{f}$ and $\hat{\mathbf{w}}$ on $\hat{W}$:

$$\nabla f = \hat{\mathbf{F}}_W^{-T}\hat{\nabla}\hat{f}, \quad \nabla \mathbf{w} = \hat{\nabla}\hat{\mathbf{w}}\hat{\mathbf{F}}_W^{-1}. \tag{1.17}$$

For temporal derivatives transformed to general coordinate systems $\hat{W}$ we must take care of two different velocities: the particle velocity $\hat{\mathbf{v}}$ and the domain velocity $\partial_t \hat{\mathsf{T}}_W$, which do not coincide, if $\hat{W} \neq \hat{V}$:

Lemma 1.11 (Transformation of temporal derivatives)**.** Let $f : V(t) \to \mathbb{R}$ with counterpart $\hat{f}(\hat{x}_W, t) = f(x, t)$. Given sufficient regularity, it holds

$$\partial_t f = \partial_t \hat{f} - (\hat{\mathbf{F}}_W^{-1} \partial_t \hat{\mathsf{T}}_W \cdot \hat{\nabla}) \hat{f}, \quad d_t f = \partial_t \hat{f} + (\hat{\mathbf{F}}_W^{-1} (\hat{\mathbf{v}} - \partial_t \hat{\mathsf{T}}_W) \cdot \hat{\nabla}) \hat{f}.$$

Proof. With $\hat{x}_W = \mathsf{T}_W(x, t)$ it holds

$$\partial_t f(x, t) = d_t \hat{f}(\hat{x}_W, t) = d_t \hat{f}(\mathsf{T}_W(x, t), t) = \partial_t \hat{f} + \hat{\nabla} \hat{f} \cdot \partial_t \mathsf{T}_W.$$

The first result follows with help of Lemma 1.10. The relation for the material derivative is given by

$$d_t f(x, t) = \partial_t f(x, t) + \nabla f \cdot \partial_t x.$$

Here, $\partial_t x = \mathbf{v} = \hat{\mathbf{v}}$ refers to the trace of particles, where $\mathbf{v} = \hat{\mathbf{v}}$ is the velocity of the particle and not the velocity of the mapping $\hat{\mathsf{T}}_W$. Together with (1.17) and the transformation of the partial time derivative we get

$$d_t f = \partial_t \hat{f} - (\hat{\mathbf{F}}_W^{-1} \partial_t \hat{\mathsf{T}}_W \cdot \hat{\nabla}) \hat{f} + \hat{\mathbf{F}}_W^{-\mathsf{T}} \hat{\nabla} \hat{f} \cdot \hat{\mathbf{v}}.$$

□

Remark 1.12 (Transformation between Lagrangian and Eulerian coordinates)**.** If $\hat{V} = \hat{W}$ it holds $\hat{\mathsf{T}}_W = \hat{\mathsf{T}}$ as well as $\hat{\mathbf{F}}_W = \hat{\mathbf{F}}$ and $\hat{J}_W = \hat{J}$. The statements of Lemma 1.11 simplify to

$$\hat{W} = \hat{V} \quad \Rightarrow \quad \partial_t f = \partial_t \hat{f} - (\hat{\mathbf{F}}^{-1} \hat{\mathbf{v}} \cdot \hat{\nabla}) \hat{f}, \quad d_t f = \partial_t \hat{f}.$$

This results explains, why the convective term $(\mathbf{v} \cdot \nabla)\mathbf{v}$ will not appear in Lagrangian coordinates. See also Lemma 1.3. △

In the following we discuss the transformation of the conservation principles to arbitrary coordinate reference systems $\hat{W}$. This transformation will be fundamental for solid mechanics, where the natural view-point is the Lagrangian one with $\hat{W} = \hat{V}$. We proceed without specifying the connotation of $\hat{W}$. The equation for conservation of momentum (1.15) is given by

$$\rho \partial_t \mathbf{v} + \rho(\mathbf{v} \cdot \nabla)\mathbf{v} = \rho \mathbf{f} + \operatorname{div}(\boldsymbol{\sigma}) \text{ in } V(t),$$

with a density ρ, velocity $\mathbf{v}$, volume force $\mathbf{f}$ and the Eulerian stress-tensor $\boldsymbol{\sigma}$. The specific form of this stress-tensor will be discussed in later sections. Here, we only assume that this stress tensor is symmetric $\boldsymbol{\sigma} = \boldsymbol{\sigma}^{\mathsf{T}}$. By $\hat{\mathbf{v}}(\hat{x}_W, t) = \mathbf{v}(x, t)$, $\hat{\rho}(\hat{x}_W, t) = \rho(x, t)$, $\hat{\mathbf{f}}(\hat{x}_W, t) = \mathbf{f}(x, t)$ as well as $\hat{\boldsymbol{\sigma}}(\hat{x}_W, t) = \boldsymbol{\sigma}(x, t)$ we denote the counterparts of these quantities in the reference system $\hat{W}$. By (1.17) and 1.11 it holds:

$$\begin{aligned} \partial_t \mathbf{v} &= \partial_t \hat{\mathbf{v}} - (\hat{\mathbf{F}}_W^{-1} \partial_t \hat{\mathsf{T}}_W \cdot \hat{\nabla}) \hat{\mathbf{v}}, \\ (\mathbf{v} \cdot \nabla)\mathbf{v} &= \nabla \mathbf{v} \mathbf{v} = \hat{\nabla} \hat{\mathbf{v}} \hat{\mathbf{F}}_W^{-1} \hat{\mathbf{v}} = (\hat{\mathbf{F}}_W^{-1} \hat{\mathbf{v}} \cdot \hat{\nabla}) \hat{\mathbf{v}}, \end{aligned}$$

and combined, we get:

$$\partial_t \mathbf{v} + (\mathbf{v} \cdot \nabla)\mathbf{v} = \partial_t \hat{\mathbf{v}} + (\hat{\mathbf{F}}_W^{-1}(\hat{\mathbf{v}} - \partial_t \hat{T}_W) \cdot \hat{\nabla})\hat{\mathbf{v}}. \tag{1.18}$$

As discussed above, in the case of a mapping to the Lagrangian reference system, the mapping's temporal derivative is the velocity $\partial_t \hat{T} = \hat{\mathbf{v}}$ and the momentum terms simplify to

$$\hat{V} = \hat{W} \quad \Rightarrow \quad \partial_t \mathbf{v} + (\mathbf{v} \cdot \nabla)\mathbf{v} = \partial_t \hat{\mathbf{v}}. \tag{1.19}$$

It remains to transform the divergence of the stresses to the reference domain. Here, a simple transformation of $\operatorname{div}(\boldsymbol{\sigma})$ to the reference system is not sufficient. We need to keep the meaning of this stress-term in mind, indicating surface-forces in normal-direction. The normal vectors are transformed, if the underlying domain $\hat{V} \to V(t)$ is deformed. Therefore we must base the mapping process on the correct representation of these surface forces. We need to find a representation of the first Piola-Kirchhoff stress tensor $\hat{\mathbf{P}}$ in the reference system, such that it holds:

$$\int_{\partial \hat{W}} \hat{\mathbf{P}}\hat{\mathbf{n}} \, d\hat{s} = \int_{\partial V(t)} \boldsymbol{\sigma}\mathbf{n} \, ds.$$

$\hat{\mathbf{P}}$ will be called the *Piola transformation* of $\boldsymbol{\sigma}$. For the derivation of this transformation we first regard vector fields $\mathbf{w} : V(t) \to \mathbb{R}^d$ with reference counterpart $\hat{\mathbf{w}} : \hat{W} \to \mathbb{R}^d$.

Lemma 1.13 (Piola transformation)**.** Let $\mathbf{w} : V(t) \to \mathbb{R}^d$ be a differentiable vector field and $\hat{\mathbf{w}}$ its representation in the reference system $\hat{W}$. The *Piola transformation* of $\mathbf{w}$ is given by

$$\hat{J}_W \hat{\mathbf{F}}_W^{-1} \hat{\mathbf{w}}.$$

On every volume $V(t)$ with corresponding reference volume $\hat{W}$ it holds

$$\int_{\partial V(t)} \mathbf{n} \cdot \mathbf{w} \, ds \int_{\partial \hat{W}} \hat{\mathbf{n}} \cdot (\hat{J}_W \hat{\mathbf{F}}_W^{-1} \hat{\mathbf{w}}) \, d\hat{s},$$
$$\int_{V(t)} \operatorname{div}(\mathbf{w}) \, dx = \int_{\hat{W}} \widehat{\operatorname{div}} (\hat{J}_W \hat{\mathbf{F}}_W^{-1} \hat{\mathbf{w}}) \, d\hat{x}.$$

Further, in a point-wise sense it holds

$$\hat{J}_W \operatorname{div}(\mathbf{w}) = \widehat{\operatorname{div}} (\hat{J}_W \hat{\mathbf{F}}_W^{-1} \hat{\mathbf{w}}).$$

Proof. We use a variational argument. Let ξ be differentiable on $V(t)$ with reference counterpart $\hat{\xi} \in \hat{W}$, such that

$$\int_{\partial V(t)} \mathbf{n} \cdot \mathbf{w}\xi \, ds = \int_{V(t)} \operatorname{div}(\mathbf{w}\xi) \, dx = \int_{\hat{W}} \hat{J}_W \operatorname{div}(\mathbf{w}\xi) \, d\hat{x}. \tag{1.20}$$

Next, with (1.17) we get for $\hat{\xi} = \xi$:

$$\int_{\hat{W}} \hat{J}_W \operatorname{div}(\mathbf{w}\xi) \, d\hat{x} = \int_{\hat{W}} \hat{J}_W \operatorname{div}(\mathbf{w}) \hat{\xi} \, d\hat{x} + \int_{\hat{W}} \hat{J}_W \hat{\mathbf{w}} \cdot \hat{\mathbf{F}}_W^{-T} \hat{\nabla} \hat{\xi} \, d\hat{x}. \tag{1.21}$$

With Green's formula, the second integral is transformed to

$$\begin{aligned}\int_{\hat{W}} \hat{J}_W \hat{\mathbf{w}} \cdot \hat{\mathbf{F}}_W^{-T} \hat{\nabla} \hat{\xi}\, d\hat{x} &= \int_{\hat{W}} \hat{J}_W \hat{\mathbf{F}}_W^{-1} \hat{\mathbf{w}} \cdot \hat{\nabla} \hat{\xi}\, d\hat{x} \\ &= -\int_{\hat{W}} \widehat{\operatorname{div}}\,(\hat{J}_W \hat{\mathbf{F}}_W^{-1} \hat{\mathbf{w}}) \hat{\xi}\, d\hat{x} + \int_{\partial \hat{W}} \hat{\mathbf{n}} \cdot (\hat{J}_W \hat{\mathbf{F}}_W^{-1} \hat{\mathbf{w}})\, \hat{\xi}\, d\hat{s}.\end{aligned} \tag{1.22}$$

Combining (1.20), (1.21) and (1.22) gives

$$\begin{aligned}\int_{\partial V(t)} \mathbf{n} \cdot \mathbf{w}\, \xi\, ds - \int_{\hat{W}} \hat{J}_W \operatorname{div}(\mathbf{w}) \hat{\xi}\, d\hat{x} \\ = -\int_{\hat{W}} \widehat{\operatorname{div}}\,(\hat{J}_W \hat{\mathbf{F}}_W^{-1} \hat{\mathbf{w}}) \hat{\xi}\, d\hat{x} + \int_{\partial \hat{W}} \hat{\mathbf{n}} \cdot (\hat{J}_W \hat{\mathbf{F}}_W^{-1} \hat{\mathbf{w}}) \hat{\xi}\, d\hat{s}.\end{aligned}$$

By picking a Dirac sequence $\{\hat{\xi}_\epsilon^y\}_{\epsilon>0}$ where $\hat{\xi}_\epsilon^y \in C_0^\infty(\hat{W})$ with

$$\int_{\hat{W}} \hat{\xi}_\epsilon^y(\hat{x}) \hat{f}(\hat{x})\, d\hat{x} \xrightarrow[\epsilon \to 0]{} \hat{f}(\hat{y}) \quad \forall \hat{f} \in C(\hat{W}),$$

we conclude for all inner points

$$\hat{J}_W \operatorname{div}(\mathbf{w}) = \widehat{\operatorname{div}}\,(\hat{J}_W \hat{\mathbf{F}}_W^{-1} \hat{\mathbf{w}}).$$

Hence

$$\int_{V(t)} \operatorname{div}(\mathbf{w})\, dx = \int_{\hat{W}} \hat{J}_W \operatorname{div}(\hat{\mathbf{w}})\, d\hat{x} = \int_{\hat{W}} \widehat{\operatorname{div}}\,(\hat{J}_W \hat{\mathbf{F}}_W^{-1} \hat{\mathbf{w}})\, d\hat{x}.$$

The relation for the surface integral follows by Gauss' divergence theorem. □

This important result is used to transform the surface forces to the reference system. Let $\boldsymbol{\sigma} = (\boldsymbol{\sigma}_i)_{i=1}^d$ be the row-vectors (or the column-vectors since $\boldsymbol{\sigma} = \boldsymbol{\sigma}^T$ by the conservation of angular momentum). It holds:

$$F_i(\partial V(t)) := \int_{\partial V(t)} \mathbf{n} \cdot \boldsymbol{\sigma}_i\, ds = \int_{V(t)} \operatorname{div}(\boldsymbol{\sigma}_i)\, dx$$

and with the just proven lemma we conclude

$$F_i(\partial V(t)) = \int_{\hat{W}} \widehat{\operatorname{div}}(\hat{J}_W \hat{\mathbf{F}}_W^{-1} \hat{\boldsymbol{\sigma}}_i)\, d\hat{x} = \int_{\partial \hat{W}} \hat{\mathbf{n}} \cdot (\hat{J}_W \hat{\mathbf{F}}_W^{-1} \hat{\boldsymbol{\sigma}}_i)\, d\hat{s}.$$

Reassembling the stress-tensor $\hat{\boldsymbol{\sigma}} = (\hat{\boldsymbol{\sigma}}_i)$ we get the reference presentation of the surface forces:

$$F(\partial V(t)) = \int_{\partial \hat{W}} (\hat{J}_W \hat{\boldsymbol{\sigma}} \hat{\mathbf{F}}_W^{-T}) \hat{\mathbf{n}}\, d\hat{s} = \int_{\hat{W}} \widehat{\operatorname{div}}\,(\hat{J}_W \hat{\boldsymbol{\sigma}} \hat{\mathbf{F}}_W^{-T})\, d\hat{x}.$$

We define

Definition 1.14 (Piola Kirchhoff stress tensors)**.** The *First Piola Kirchhoff stress tensor* given by

$$\hat{\mathbf{P}} := \hat{J}_W \hat{\boldsymbol{\sigma}} \hat{\mathbf{F}}_W^{-T}.$$

It relates forces in the Eulerian coordinate framework with coordinates in a reference framework $\hat{W}$. The *Second Piola Kirchhoff stress tensor* given by

$$\hat{\boldsymbol{\Sigma}} := \hat{\mathbf{F}}_W^{-1} \hat{\mathbf{P}} = \hat{J}_W \hat{\mathbf{F}}_W^{-1} \hat{\boldsymbol{\sigma}} \hat{\mathbf{F}}_W^{-T}.$$

Unlike the Eulerian stress tensor $\boldsymbol{\sigma}$, the 1st Piola Kirchhoff stress tensor $\hat{\mathbf{P}}$ is not symmetric. The 2nd Piola Kirchhoff stress tensor is symmetric but it does not have an immediate physical explanation.

Using the first Piola Kirchhoff stress tensor and Relation (1.18) the momentum equation on arbitrary reference systems $\hat{W}$ is given by:

$$\hat{J}_W \hat{\rho}\big(\partial_t \hat{\mathbf{v}} + (\hat{\mathbf{F}}_W^{-1}(\hat{\mathbf{v}} - \partial_t \hat{T}_W) \cdot \hat{\nabla})\hat{\mathbf{v}}\big) = \hat{J}_W \hat{\rho} \hat{\mathbf{f}} + \widehat{\operatorname{div}}\,(\hat{J}_W \hat{\boldsymbol{\sigma}} \hat{\mathbf{F}}_W^{-T}). \tag{1.23}$$

1.2 Material laws

The basic concepts of continuum mechanics introduced in the previous section are exact in a way that they are based on fundamental physical principles. The conservation principles for mass, momentum and angular momentum constitute a systems of four partial differential equations for 10 unknowns: density ρ, velocity field $\vec{v}$ and the six unknowns of the symmetric stress tensor $\boldsymbol{\sigma}$. This system is under-determined. To close it additional equations are required that connect the values of the stress tensor to computable fundamental quantities like velocity, density or deformation.

In the following sections, we will derive such *material laws* that describe the properties of the stress tensors in the different formulations like $\boldsymbol{\sigma}$, $\hat{\boldsymbol{\Sigma}}$ or $\hat{\mathbf{P}}$. We assume that these stress tensors will depend on strain or strain rate given as deformation gradient $\hat{\mathbf{F}}$, its inverse $\mathbf{F}$, or tensors like $\hat{\mathbf{C}}$, $\hat{\mathbf{E}}$, $\mathbf{b}$, $\mathbf{e}$ or $\dot{\boldsymbol{\epsilon}}$. We denote this relation by tensor-valued functions

$$\boldsymbol{\sigma} = f(\dot{\boldsymbol{\epsilon}}), \quad \hat{\mathbf{P}} = \hat{f}(\hat{\mathbf{F}}), \quad \hat{\boldsymbol{\Sigma}} = \hat{f}(\hat{\mathbf{E}}),$$

or by similar expressions in $\hat{\mathbf{E}}$ or $\mathbf{b}$. We assume that all materials are homogenous and do not explicitly depend on the location $x \in V(t)$.

We are not considering arbitrary material laws but postulate several assumption on the material's properties:

1. *Objectivity:* The material law is independent of the spectators viewpoint. This property will hold for every physical material.

2. *Homogeneity:* We assume that the material is homogenous, i.e. the strain-stress relation will not explicitly depend on the location $x \in V(t)$.

3. *Isentropic and isothermal processes:* We assume that entropy and temperature do not play a role. There is no conversion between heat and kinetic energy. The temperature stays constant and does not affect the material law. This assumption is a simplification, as most elastic materials and also some fluids show a strong dependency on the temperature.

4. *Isotropy:* There is no distinct direction in the material. The response to strain or strain rate is the same in all directions. This assumption rules out anisotropic materials like fiber-reinforced composites or also biological tissue, where layers are usually directed anisotropically. Most fluids however are isotropic.

These assumptions lead to a strong simplification of possible material laws. The following *Rivlin-Ericksen Theorem* shows that all such possible material laws depend on symmetric strain tensors $\mathbf{C}$, $\mathbf{E}$ or $\dot{\boldsymbol{\epsilon}}$ only and that all material laws are quadratic polynomials in the invariants of these tensors:

Theorem 1.15 (Rivlin-Ericksen Theorem)**.** A stress response function $\tilde{f}(\hat{\mathbf{F}})$ is isotropic and indifferent with respect to the coordinate system, if and only if it depends on the symmetric strain tensors only

$$\tilde{f}(\hat{\mathbf{F}}) = \hat{f}(\hat{\mathbf{F}}^{\mathsf{T}}\hat{\mathbf{F}}) = \hat{f}(\hat{\mathbf{C}}).$$

Further it is given as a quadratic polynomial

$$\hat{f}(\hat{\mathbf{C}}) = \beta_0(i(\hat{\mathbf{C}}))\mathrm{I} + \beta_1(i(\hat{\mathbf{C}}))\hat{\mathbf{C}} + \beta_2(i(\hat{\mathbf{C}}))\hat{\mathbf{C}}^2, \tag{1.24}$$

with scalar coefficients β_i that depend on the invariants (under orthogonal transformation) of the symmetric tensors $\mathbf{C}$:

$$\mathrm{I}_1(\vec{\mathbf{C}}) = \lambda_1 + \lambda_2 + \lambda_3, \quad \mathrm{I}_2(\vec{\mathbf{C}}) = \lambda_1\lambda_2 + \lambda_2\lambda_3 + \lambda_1\lambda_3, \quad \mathrm{I}_3(\vec{\mathbf{C}}) = \lambda_1\lambda_2\lambda_3,$$

where λ_1, λ_2 and λ_3 are the three eigenvalues of $\vec{\mathbf{C}}$.

Proof. For a proof, we refer to the original contribution by Rivlin and Ericksen [32] or to a modern presentation by Turesdell and Noll [39]. □

As a symmetric positive definite tensor, $\mathbf{C}$ has three positive eigenvalues $\lambda_1, \lambda_2, \lambda_3$ and a system of orthogonal eigenvectors. We know that eigenvalues are invariant to orthogonal transformation. To derive these invariants, we further cite the following Lemma:

Lemma 1.16. Given a tensor $\vec{A} \in \mathbb{R}^{3\times 3}$ it holds for every $\lambda \in \mathbb{R}$

$$\det(\vec{A} - \lambda\mathrm{I}) = -\lambda^3 + \mathrm{I}_1(\vec{A})\lambda^2 + \mathrm{I}_2(\vec{A})\lambda + \mathrm{I}_3(\vec{A}),$$

with

$$\mathrm{I}_1(\vec{A}) = \mathrm{tr}(\vec{A}), \quad \mathrm{I}_2(\vec{A}) = \frac{1}{2}\left(\mathrm{tr}(\vec{A})^2 - \mathrm{tr}(\vec{A}^2)\right), \quad \mathrm{I}_3(\vec{A}) = \det(\vec{A}).$$

If $\vec{A}$ is symmetric positive definite with eigenvalues $\lambda_1, \lambda_2, \lambda_3$, it further holds

$$\mathrm{I}_1(\vec{A}) = \lambda_1 + \lambda_2 + \lambda_3, \quad \mathrm{I}_2(\vec{A}) = \lambda_1\lambda_2 + \lambda_2\lambda_3 + \lambda_1\lambda_3, \quad \mathrm{I}_3(\vec{A}) = \lambda_1\lambda_2\lambda_3.$$

Proof. See [21]. □

The Rivlin-Ericksen Theorem 1.15 strongly limits possible material laws for homogenous and isotropic materials. All material laws - including fluids and solids - considered in the context of this book will fall under this theorem.

As every matrix satisfies its own characteristic polynomial, it holds for $\hat{\mathbf{C}} \in \mathbb{R}^{3\times 3}$ that

$$\hat{\mathbf{C}}^3 = \mathrm{I}_1(\hat{\mathbf{C}})\hat{\mathbf{C}}^2 + \mathrm{I}_2(\hat{\mathbf{C}})\hat{\mathbf{C}} + \mathrm{I}_3(\hat{\mathbf{C}}). \tag{1.25}$$

Using this relation, the material law (1.24) is equivalent to a second representation

$$\hat{\mathrm{f}}(\hat{\mathbf{C}}) = \gamma_0(\mathfrak{i}(\hat{\mathbf{C}}))\mathrm{I} + \gamma_1(\mathfrak{i}(\hat{\mathbf{C}}))\hat{\mathbf{C}} + \gamma_2(\mathfrak{i}(\hat{\mathbf{C}}))\hat{\mathbf{C}}^{-1}.$$

Remark 1.17. As the two tensors $\hat{\mathbf{E}} = \frac{1}{2}(\hat{\mathbf{C}} - \mathrm{I})$ are directly connected, every material law in $\hat{\mathbf{C}}$ can also be expressed in $\hat{\mathbf{E}}$, as

$$\alpha_0\mathrm{I} + \alpha_1\hat{\mathbf{C}} + \alpha_2\hat{\mathbf{C}}^2 = (\alpha_0 + \alpha_1 + \alpha_2)\mathrm{I} + (2\alpha_1 + 4\alpha_2)\hat{\mathbf{E}} + 4\alpha_2\hat{\mathbf{E}}^2.$$

Further, for the eigenvalues of $\hat{\mathbf{E}}$ and $\hat{\mathbf{C}}$ there holds a linear relation

$$\lambda_\mathfrak{i} w_\mathfrak{i} = \hat{\mathbf{C}} w_\mathfrak{i} = 2\hat{\mathbf{E}} w_\mathfrak{i} + w_\mathfrak{i} \quad \Leftrightarrow \quad \frac{1}{2}(\lambda_\mathfrak{i} - 1) w_\mathfrak{i} = \hat{\mathbf{E}} w_\mathfrak{i}.$$

△

1.2.1 Hyperelastic materials

A solid is called *hyperelastic* if the relation between strain and stress comes from an energy density function

$$\hat{\mathbf{\Sigma}} = \frac{\partial W(\hat{\mathbf{E}})}{\partial \hat{\mathbf{E}}},$$

or

$$\hat{\mathbf{P}} = \frac{\partial W(\hat{\mathbf{F}})}{\partial \hat{\mathbf{F}}}.$$

This constitutes a relation between the second Piola-Kirchhoff stress tensor and the strain or between the deformation gradient and the first Piola-Kirchhoff stress, respectively. Many of the commonly used materials like the *St. Venant Kirchhoff* model or the *Mooney-Rivlin solid* are of this type. Stress tensors for incompressible materials can be derived by energy functions of the type

$$W = W(\mathbf{F}) - \mathfrak{p}\,(\det(\mathbf{F}) - 1)$$

that penalize the change of volume $\mathrm{J} = \det(\mathbf{F})$.

As the derivation of the models is not in the focus of this book, we just refer to the literature for more reading on this very important concept, see Holzapfel [21] for a very comprehensive exposure.

1.2.2 Linearizations

For simplicity, we sometimes consider linear models. Two different types of nonlinearities must be considered: first, the material nonlinearity which denotes a nonlinear relation between stress and strain. Second, the geometric nonlinearity, which comes from the discrepancy between reference coordinate system and current system and which is expressed by the deformation gradient $\mathbf{e} = \hat{\mathbf{F}}\hat{\mathbf{e}}$.

Regarding the Rivlin-Ericksen Theorem 1.15, linearity of a material means that only the first invariant $I_1(\mathbf{E}) = \operatorname{tr}(\mathbf{E})$ may enter the law and that no higher order terms may appear. Further, in geometrically linearized situations, the symmetric strain tensor and $\hat{\mathbf{E}}$ is approximated and linearized

$$\hat{\mathbf{E}} = \frac{1}{2}(\nabla\mathbf{u} + \nabla\mathbf{u}^T + \nabla\mathbf{u}^T\nabla\mathbf{u}) \approx \frac{1}{2}(\nabla\mathbf{u} + \nabla\mathbf{u}^T) =: \hat{\boldsymbol{\epsilon}},$$

assuming that $|\nabla\mathbf{u}| \ll 1$ is small.

Lemma 1.18 (Linear material law)**.** A stress response function $f(\cdot)$ for a linear, homogenous and isotropic material depends on the linearized strain $\hat{\boldsymbol{\epsilon}} = \hat{\nabla}\hat{\mathbf{u}} + \hat{\nabla}\hat{\mathbf{u}}^T$ or on the strain rate tensor $\dot{\boldsymbol{\epsilon}} = \nabla\mathbf{v} + \nabla\mathbf{v}^T$ and its first invariant only

$$\hat{f}(\hat{\boldsymbol{\epsilon}}) = \beta_0 \operatorname{tr}(\hat{\boldsymbol{\epsilon}})I + \beta_1\hat{\boldsymbol{\epsilon}}.$$

In fluid mechanics, the Navier-Stokes equations follow such a linear material law and in structure mechanics, the Navier-Lamé problem considers these simplifications. While in fluid mechanics a fully linear material law - the Navier-Stokes model - is a very accurate model for many relevant fluids, linearization in solid mechanics is usually not feasible. Here, linear models only apply to very small deformations $|\hat{\mathbf{u}}| \ll 1$ and very small changes in deformation $|\hat{\nabla}\hat{\mathbf{u}}| \ll 1$. In particular, linearized solid models are no longer invariant with respect to fixed body rotations.

1.2.3 Incompressible materials

Some materials have an incompressible behavior which means that the volume

$$|V(t)| = \int_{V(t)} 1\,dx \equiv \text{“const”}$$

does not change. For an incompressible material, there is no expansion or compression. Many fluids - like water - can be considered incompressible. Incompressibility further applies to many biological structures. We can describe change of volume in the current system by Reynolds transport theorem

$$0 = d_t|V(t)| = d_t \int_{V(t)} 1\,dx = \int_{V(t)} \nabla\cdot\mathbf{v}\,dx = \int_{\partial V(t)} n\cdot\mathbf{v}\,ds, \tag{1.26}$$

but also in the reference configuration by transformation

$$0 = d_t|V(t)| = d_t \int_{V(t)} 1 \, dx = \int_{\hat{V}} d_t \hat{J} \, d\hat{x}. \tag{1.27}$$

For a fluid, modeled in the current configuration, (1.26) says that the flow is "divergence-free" with $\operatorname{div} \mathbf{v} = 0$ and also that the total normal flow over the volume's boundary is zero. For a divergence free velocity field it holds

$$\operatorname{tr}(\dot{\boldsymbol{\varepsilon}}) = 0,$$

and in light of Lemma 1.18, the material law is further simplified to

$$f(\dot{\boldsymbol{\varepsilon}}) = \beta_1 \dot{\boldsymbol{\varepsilon}}.$$

To cope with isotropic expansion and compression forces, we introduce a pressure variable as part of the material law:

$$f(\dot{\boldsymbol{\varepsilon}}, p) = -pI + \beta_1 \dot{\boldsymbol{\varepsilon}}.$$

This pressure will be required to enforce the incompressibility of the velocity field.

Considering solid's, incompressibility in terms of (1.27) means that the determinant of the deformation gradient will be constant $d_t \hat{J} = 0$. As $\hat{\mathbf{F}} = I$ in the reference system, incompressibility simply says $\hat{J} = 1$ for all times $t \geqslant t_0$. Further, it then holds that

$$\det(\hat{\mathbf{C}}) = \det(\hat{\mathbf{F}})^2 = 1.$$

For the Green-Lagrange strain tensor $\hat{\mathbf{E}}$ it follows that third and second invariant fall together, see Lemma 1.16 and Remark 1.17.

1.3 The solid problem

As discussed, we usually describe the dynamics of elastic structures in the Lagrangian reference system. Hence considering the conservation law (1.23) we choose $\hat{W} = \hat{V}$ as reference system. In light of Remark 1.12, the momentum equation is given by

$$\hat{J}\hat{\rho}\partial_{tt}\hat{\mathbf{u}} = \hat{J}\hat{\rho}\hat{\mathbf{f}} + \widehat{\operatorname{div}}(\hat{\mathbf{F}}\hat{\boldsymbol{\Sigma}}),$$

where we eliminated the velocity using $\partial_t \hat{\mathbf{u}} = \hat{\mathbf{v}}$. Considering material laws as introduced in the previous section, stresses will depend on strain, and hence on the displacement $\hat{\mathbf{u}}$. The density is known at initial time $\rho(x, 0) = \hat{\rho}^0(\hat{x})$. For $t \geqslant 0$ conservation of mass yields

$$m(\hat{V}) := \int_{\hat{V}} \hat{\rho}^0(\hat{x}) \, d\hat{x} \overset{!}{=} \int_{V(t)} \rho(x, t) \, dx = \int_{\hat{V}} \hat{J}\hat{\rho}(\hat{x}, t) \, d\hat{x} =: m(V(t)).$$

At time $t \geqslant 0$, the relation

$$\hat{\rho}(\hat{x}, t) = \hat{J}^{-1}(\hat{x}, t)\hat{\rho}^0(\hat{x}) \tag{1.28}$$

describes the density in every point $\hat{x}$ of the reference system. The full problem of elastic structures formulated in the Lagrangian reference system $\hat{V}$ is given by:

$$\hat{\rho}^0 \partial_{tt} \hat{\mathbf{u}} - \widehat{\operatorname{div}}\,(\mathbf{F}\hat{\mathbf{\Sigma}}) = \hat{\rho}^0 \hat{\mathbf{f}} \tag{1.29}$$

It remains to complete this partial differential equation by appropriate boundary conditions and initial conditions. Let $\hat{S} \subset \mathbb{R}^d$ be the solid domain in reference configuration. At time $t = 0$, we specify initial conditions for density, deformation and velocity

$$\hat{\rho}(\cdot, 0) = \hat{\rho}^0(\cdot), \quad \hat{\mathbf{u}}(\cdot, 0) = \hat{\mathbf{u}}^0(\cdot), \quad \partial_t \hat{\mathbf{u}}(\cdot, 0) = \hat{\mathbf{v}}^0(\cdot), \quad t = 0. \tag{1.30}$$

For all times $t \geqslant 0$, by $\hat{\mathbf{f}}(\hat{x}, t)$ we denote the acting volume force field. Note that this force field is directed in the Eulerian framework, such that for example the gravity is given by $\mathbf{f} = -9.81 e_3 \mathrm{kg} \cdot \mathrm{m} \cdot \mathrm{s}^{-2}$, with $e_3 = (0, 0, 1)^T$, independent of the reference framework. The boundary of the domain $\hat{\Gamma}_s := \partial\hat{S}$ is split into a Dirichlet boundary part $\hat{\Gamma}_s^D$ and into a Neumann part $\hat{\Gamma}_s^N$. On the Dirichlet boundary, we specify boundary conditions for the deformation

$$\hat{\mathbf{u}} = \hat{\mathbf{u}}^D \quad \text{ on } \hat{\Gamma}_s^D \times [0, T]. \tag{1.31}$$

Note that by $\hat{\mathbf{v}} = \partial_t \hat{\mathbf{u}}$ we also uniquely define the velocity on the boundary. The usual Neumann condition on Γ_s^N specifies the boundary stresses by

$$\mathbf{n} \cdot \hat{\mathbf{F}}\hat{\mathbf{\Sigma}} = \mathbf{n} \cdot \hat{J}\hat{\sigma}_s \hat{\mathbf{F}}^{-T} = \hat{\mathbf{g}}_s^{(\hat{n})} \text{ on } \hat{\Gamma}_s^N \times [0, T]. \tag{1.32}$$

If the external forces $\mathbf{f}$ and the boundary data $\hat{\mathbf{g}}_s^{(\hat{n})}$ and $\hat{\mathbf{u}}^D$ do not explicitly depend on time, the solution can run into a stationary limit $\hat{\mathbf{u}}(\cdot, t) \to \hat{\mathbf{u}}(\cdot)$ that does not depend on time. In this case, it holds $\partial_t \hat{\mathbf{v}} = 0$ and hence $\partial_{tt} \hat{\mathbf{u}} = 0$. If such a stationary solution exists, we can directly consider the stationary system of equations:

$$-\widehat{\operatorname{div}}\,(\hat{\mathbf{F}}\hat{\mathbf{\Sigma}}) = \hat{\rho}^0 \hat{\mathbf{f}}. \tag{1.33}$$

Finally, it remains to provide material laws for specific solids. One of the most simple model is the *St. Venant Kirchhoff material* that postulates a linear dependency between strain tensor $\hat{\mathbf{E}}$ and stresses:

Definition 1.19 (St. Venant Kirchhoff material)**.** The St. Venant Kirchhoff material follows the material law

$$\hat{\mathbf{\Sigma}} = 2\mu_s \hat{\mathbf{E}} + \lambda_s \operatorname{tr}(\hat{\mathbf{E}}) I,$$

with the first λ_s and second μ_s Lamé parameters. (μ_s is also called the *shear modulus*.) These two parameters are related to the Poisson ratio ν_s that describes the compressibility and Young's modulus E_s that describes the stiffness:

$$\nu_s = \frac{\lambda_s}{2(\lambda_s + \mu_s)}, \quad E_s = \frac{\mu_s(3\lambda_s + 2\mu_s)}{\lambda_s + \mu_s}.$$

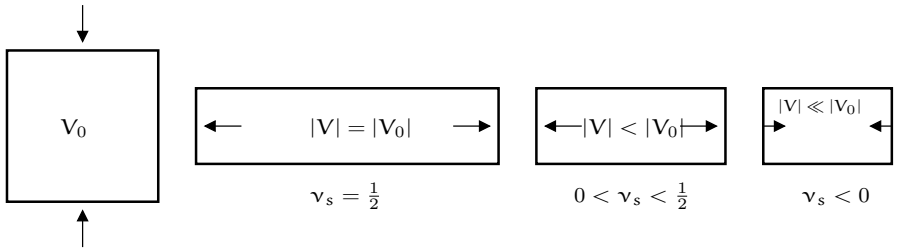


Figure 1.6: Material behavior under compression for different Poisson ratios. Left: incompressible material $\nu_s = \frac{1}{2}$. Middle: compressible material $0 < \nu_s < \frac{1}{2}$. Right: auxetic material with $\nu_s < 0$.

The linear relation between strain and stress is called *Hooke's Law*. The Poisson ratio ν_s describes the compressibility of the system. It holds

$$\nu_s = \frac{1}{2}\left(\frac{1}{1+\frac{\mu_s}{\lambda_s}}\right) < \frac{1}{2}.$$

The Poisson ratio $\nu_s = \frac{1}{2}$ refers to $\lambda_s \to \infty$ hence to incompressible materials. The Poisson ratio describes the reaction of the material on directional compression, see Figure 1.6. For a Poisson ratio $\nu_s = \frac{1}{2}$, the volume will stay constant, for $\nu_s < \frac{1}{2}$ the volume will decrease. There are some materials with negative Poisson ratio. Here, the material will react to the compression in one direction with compression in the orthogonal directions. The St. Venant Kirchhoff model is a suitable approximation for metals at small deformations. Steel has a Poisson ratio of about $\nu_s \approx 0.3$ and a Young modulus $E_s \approx 200 \cdot 10^9 \mathrm{kg} \cdot \mathrm{m}^{-1} \cdot \mathrm{s}^{-2}$.

Hooke's Law applied to an incompressible material leads to the incompressible Neo Hookean material law:

Definition 1.20 (Incompressible Neo-Hookean material)**.** The incompressible Neo-Hookean material law is given by

$$\hat{\mathbf{P}} = \hat{\mathbf{F}}\hat{\mathbf{\Sigma}} = -p\hat{\mathbf{F}}^{-T} + 2\mu_s\hat{\mathbf{F}}^{-T}\hat{\mathbf{E}},$$

with the shear modulus μ_s and the Poisson ratio $\nu_s = \frac{1}{2}$. By p we denote the undetermined pressure.

We conclude and formulate the following often used systems of equations

System 1.21 (Conservation laws for a St. Venant Kirchhoff material)**.** Let $\Omega \subset \mathbb{R}^d$ be a domain with boundary $\Gamma = \partial\Omega$ with $\Gamma = \Gamma^D \cup \Gamma^N$. Further, let $\hat{\rho}^0 : \Omega \to \mathbb{R}_+$ be the materials density, $\hat{\mathbf{f}} \in C(\Omega)^d$ be a given right hand side, $\hat{\mathbf{u}}^D, \hat{\mathbf{v}}^D \in C(\Gamma^D)$ be Dirichlet boundary data, $\hat{\mathbf{g}}^{(\mathbf{n})} \in C(\Gamma^N)$ be the Neumann data. With initial deformation and velocity $\hat{\mathbf{u}}^0, \hat{\mathbf{v}}^0 \in C(\Omega)^d$ find deformation and velocity

$$\hat{\mathbf{u}}(t) \in C^2(\Omega)^d \cap C(\Omega \cup \Gamma^D)^d \cup C^1(\Omega \cup \Gamma^N)^d,$$

such that

$$\hat{\rho}^0 \partial_{tt}\hat{\mathbf{u}} - \widehat{\operatorname{div}}\left(\hat{\mathbf{F}}\hat{\mathbf{\Sigma}}\right) = \hat{\rho}^0\hat{\mathbf{f}} \quad t \geqslant 0,$$

where

$$\hat{\mathbf{\Sigma}} = 2\mu_s\hat{\mathbf{E}} + \lambda_s \operatorname{tr}(\hat{\mathbf{E}})\mathrm{I},$$

and

$$\hat{\mathbf{u}}(0) = \hat{\mathbf{u}}^0,\ d_t\hat{\mathbf{u}}(0) = \hat{\mathbf{v}}^0 \text{ in } \Omega,$$

with the boundary conditions

$$\hat{\mathbf{u}}(t) = \hat{\mathbf{u}}^D \text{ on } \Gamma^N, \quad \hat{\mathbf{F}}\hat{\mathbf{\Sigma}}\hat{\mathbf{n}} = \hat{\mathbf{g}}^{(\mathbf{n})}.$$

For incompressible materials we define:

System 1.22 (Conservation laws for the incompressible Neo-Hookean material)**.** Let $\Omega \subset \mathbb{R}^d$ be a domain with boundary $\Gamma = \partial\Omega$ with $\Gamma = \Gamma^D \cup \Gamma^N$. Further, let $\hat{\rho}^0 : \Omega \to \mathbb{R}_+$ be the materials density, $\hat{\mathbf{f}} \in C(\Omega)^d$ be a given right hand side, $\hat{\mathbf{u}}^D, \hat{\mathbf{v}}^D \in C(\Gamma^D)$ be Dirichlet boundary data, $\hat{\mathbf{g}}^{(\mathbf{n})} \in C(\Gamma^N)$ be the Neumann data. With initial deformation and velocity $\hat{\mathbf{u}}^0, \hat{\mathbf{v}}^0 \in C(\Omega)^d$ find deformation, velocity and pressure

$$\hat{\mathbf{u}}(t) \in C^2(\Omega)^d \cap C(\Omega \cup \Gamma^D)^d \cap C^1(\Omega \cup \Gamma^N)^d, \quad \hat{p}(t) \in C^1(\Omega) \cap C(\Omega \cup \Gamma^N),$$

such that

$$\hat{J} = 0, \quad \hat{\rho}^0 \partial_{tt}\hat{\mathbf{u}} - \widehat{\operatorname{div}}\left(\hat{\mathbf{F}}\hat{\mathbf{\Sigma}}\right) = \hat{\rho}^0\hat{\mathbf{f}} \quad t \geqslant 0,$$

where

$$\hat{\mathbf{\Sigma}} = -\hat{p}\hat{\mathbf{F}}^{-T} + 2\mu_s\hat{\mathbf{F}}^{-T}\hat{\mathbf{E}}$$

and

$$\hat{\mathbf{u}}(0) = \hat{\mathbf{u}}^0,\ d_t\hat{\mathbf{u}}(0) = \hat{\mathbf{v}}^0 \text{ in } \Omega,$$

with the boundary conditions

$$\hat{\mathbf{u}}(t) = \hat{\mathbf{u}}^D \text{ on } \Gamma^N, \quad \hat{\mathbf{F}}\hat{\mathbf{\Sigma}}\hat{\mathbf{n}} = \hat{\mathbf{g}}^{(\mathbf{n})}.$$

1.3.1 The Navier-Lamé equations

The model for an elastic solid governed by one of the material laws is a system of nonlinear partial differential equations. Its analysis is difficult and theoretical results exist for small deformation only. As a nonlinear set of equations, uniqueness cannot be expected in the general case.

To get better insight into the problem, we will simplify the problem with the following assumptions:

- The deformation gradient $\hat{\mathbf{F}}$ is so small that we can approximate $\hat{\mathbf{F}} = \mathrm{I}$ and $\hat{J} = 1$. By this simplification, the concept of Eulerian and Lagrangian coordinates fall together. We will therefore also skip all hat's that indicate reference variables.

- Further the strains are so small that we can linearize the Green-Lagrange strain tensor

$$\hat{\mathbf{E}} = \frac{1}{2}(\hat{\nabla}\mathbf{u} + \hat{\nabla}\mathbf{u}^T + \hat{\nabla}\hat{\mathbf{u}}^T\hat{\nabla}\hat{\mathbf{u}}) \approx \frac{1}{2}(\nabla\mathbf{u} + \nabla\mathbf{u}) =: \boldsymbol{\epsilon}.$$

  This simplification not only rules out very large elastic deformations, it also penalizes rigid body rotations.

- Just for simplicity (this will not change the character of the equation) we set $\hat{\rho}^0 = 1$.

Considering the linear St. Venant Kirchhoff material (with these simplifications) the resulting set of equations are the

System 1.23 (Navier-Lamé equations)**.** Let $\Omega \subset \mathbb{R}^3$ be a bounded domain with a boundary split into Dirichlet- and Neumann-part $\partial\Omega = \Gamma^D \cup \Gamma^N$. On the time interval $I = [0, T]$ we search for solutions $\mathbf{u} : I \times \Omega \to \mathbb{R}^3$ such that

$$\begin{aligned} \partial_{tt}\mathbf{u} - \operatorname{div}\boldsymbol{\sigma} &= \mathbf{f} && \text{in } I \times \Omega \\ \mathbf{u} = \mathbf{u}^0, \quad d_t\mathbf{u} &= \mathbf{v}^0 && \text{for } \{0\} \times \Omega \\ \mathbf{u} &= \mathbf{u}^D && \text{on } I \times \Gamma^D \\ \boldsymbol{\sigma}\,\mathbf{n} &= \mathbf{u}^\sigma && \text{on } I \times \Gamma^N, \end{aligned} \tag{1.34}$$

with the linearized material law

$$\boldsymbol{\sigma} = 2\mu\boldsymbol{\epsilon} + \lambda\,\mathrm{tr}(\boldsymbol{\epsilon})I, \quad \boldsymbol{\epsilon} = \frac{1}{2}\left(\nabla\mathbf{u} + \nabla\mathbf{u}^T\right).$$

As a further simplification, we also consider the stationary limit of the Navier-Lamé equations:

System 1.24 (Stationary Navier-Lamé equations)**.** Find $u \in C^2(\Omega)^3 \cap C(\Omega \cup \Gamma^D)^3 \cap C^1(\Omega \cup \Gamma^N)^3$ such that

$$\begin{aligned} -\operatorname{div}\boldsymbol{\sigma} &= \mathbf{f} && \text{in } \Omega \\ \mathbf{u} &= \mathbf{u}^D && \text{on } \Gamma^D \\ \boldsymbol{\sigma}\,\mathbf{n} &= \mathbf{u}^\sigma && \text{on } \Gamma^N, \end{aligned} \tag{1.35}$$

with the linearized material law

$$\boldsymbol{\sigma} = 2\mu\boldsymbol{\epsilon} + \lambda\,\mathrm{tr}(\boldsymbol{\epsilon})I, \quad \boldsymbol{\epsilon} = \frac{1}{2}\left(\nabla\mathbf{u} + \nabla\mathbf{u}^T\right).$$

As usual, analysis of classical solutions is difficult. This is partly to the fact that the solution $\mathbf{u}$ often exhibits singularities in boundary nodes at the transit between Dirichlet and Neumann parts. The well known *Theorem of Cosserat* states that classical solutions to the stationary problem, Problem 1.24, are unique if the Dirichlet boundary Γ^D contains at least three independent points and that–in the general case–they can differ by a rigid body motion only

$$\mathbf{u}_1(x) - \mathbf{u}_2(x) = \mathbf{b} + \mathbf{B}x,$$

where $\mathbf{b} \in \mathbb{R}^3$ is a translation vector and $\mathbf{B} \in \mathbb{R}^{3\times 3}$ is a skew-symmetric matrix, see e.g. [7].

For the following, we will introduce a weak formulation of the Navier-Lamé equations that will offer an easy access to show existence and uniqueness of solutions:

Lemma 1.25 (Variational formulation)**.** Every classical solution to Problem 1.24 is also solution to the variational formulation

$$\mathbf{u} \in \bar{\mathbf{u}}^D + H_0^1(\Omega;\Gamma^D)^3$$
$$(\boldsymbol{\sigma}, \nabla\phi) = (\mathbf{f},\phi) + \langle \mathbf{u}^\sigma, \phi\rangle_{\Gamma^N} \quad \forall \phi \in H_0^1(\Omega;\Gamma^D)^3, \quad (1.36)$$

where $\bar{\mathbf{u}}^D \in H^1(\Omega)^d$ is an extension of the Dirichlet data $\mathbf{u}^D$ into the domain.

Existence and uniqueness of solutions can be shown by standard arguments of elliptic equations. The difficulty however is to show ellipticity, i.e.

$$\mu(\nabla\mathbf{u} + \nabla\mathbf{u}^T, \nabla\mathbf{u}) + \lambda\big(\operatorname{tr}(\nabla\mathbf{u} + \nabla\mathbf{u}^T)I, \nabla\mathbf{u}\big) \geqslant c\|\nabla\mathbf{u}\|^2,$$

as $\nabla\mathbf{u} + \nabla\mathbf{u}^T = 0$ does not necessarily impose $\nabla\mathbf{u} = 0$. This is a consequence of *Korn's inequality*:

Theorem 1.26 (1st Korn's inequality)**.** Let $\Omega \subset \mathbb{R}^3$ be a domain. Then, it holds

$$\|\nabla\mathbf{v}\| \leqslant c_{\text{korn}}\|\boldsymbol{\epsilon}(\mathbf{v})\| \quad \forall \mathbf{v} \in H_0^1(\Omega)^3$$

with a constant $c_{\text{korn}} > 0$. This inequality corresponds to the case of Dirichlet boundary values on the complete boundary $\Gamma^D = \partial\Omega$.

Korn's first inequality deals with the case of homogenous Dirichlet conditions on the complete boundary $\partial\Omega$. In the context of structural mechanics, this limitation is severe, as no free boundary motion and deformation would be allowed. The case of general boundary conditions, with a Neumann part $\Gamma_N \subset \partial\Omega$ is less trivial and handled by Korn's second inequality:

Theorem 1.27 (2nd Korn's inequality)**.** Let $\Omega \subset \mathbb{R}^3$ be a domain with Lipschitz-boundary. Then, it holds

$$\|\nabla\mathbf{v}\| \leqslant c_{\text{korn}}\left(\|\epsilon(\mathbf{v})\| + \|\mathbf{v}\|\right) \quad \forall \mathbf{v} \in H^1(\Omega)^3.$$

with a constant $c_{\text{korn}} > 0$.

Proof. The simple proof of 1st Korn's inequality is based on integration by parts and vanishing traces of $\mathbf{v}$ on the complete boundary $\partial\Omega$. The proof of Korn's 2nd inequality is more involved and we refer to the literature, see e.g. [22, 8]. □

Continuity and ellipticity of the bilinear form allows to apply the standard theory for linear elliptic problems to the Navier-Lamé equations.

Lemma 1.28 (Existence of unique solutions)**.** Let $\mathbf{f} \in L^2(\Omega)^3$, $\bar{\mathbf{u}}^D \in H^1(\Omega)^3$ be an extension of the Dirichlet data into the domain and $\mathbf{u}^\sigma \in H^1(\partial\Omega)^3$. There exists a unique solution $\mathbf{u} \in \bar{\mathbf{u}}^D + H_0^1(\Omega; \Gamma^D)^3$ to the linear Navier-Lamé equations and it holds

$$\|\mathbf{u}\|_{H^1(\Omega)} \leqslant c \left(\|\mathbf{f}\|_{L^2(\Omega)} + \|\mathbf{u}^D\|_{L^2(\Gamma^D)} + \|\mathbf{u}^\sigma\|_{H^1(\Gamma^N)} \right),$$

with a constant $c > 0$.

Proof. We must show that the variational formulation is bilinear, symmetric, continuous and elliptic. Further, the right hand side is continuous, such that existence of a unique solution follows by the Theorem of Lax-Milgram, see [33]. □

Concerning the regularity of the solution, we cite the following lemma, see [7], which gives conditions that lead to classical solutions.

Lemma 1.29 (Strong regularity of the Navier-Lamé problem)**.** Let $\Omega \subset \mathbb{R}^3$ be a bounded domain of class $C^{2+\alpha}$ for $\alpha > 0$. Given that the problem data has the regularity

$$\mathbf{f} \in C^\alpha(\bar{\Omega})^3, \quad \bar{\mathbf{u}}^\sigma \in C^{1+\alpha}(\Omega)^{3\times 3}_{\text{sym}}, \quad \bar{\mathbf{u}}^D \in C^{2+\alpha}(\bar{\Omega})^3,$$

the weak solution $\mathbf{u} \in H_0^1(\Omega; \Gamma^D)^3$ of (1.36) is also a classical solution

$$\mathbf{u} \in C^2(\Omega)^3 \cap C^1(\Omega \cup \Gamma^N)^3 \cap C(\Omega \cup \Gamma^D)^3.$$

A further regularity result with less strict assumption on the regularity of the domain and the problem data is given by Shi and Wright [36]:

Lemma 1.30 (Weak regularity of the Navier-Lamé problem)**.** Let $\Omega \subset \mathbb{R}^3$ be a domain with $W^{2,3}$ boundary. Further, let $\mathbf{f} \in L^2(\Omega)^d$. Then, for the solution of the stationary Navier-Lamé problem with homogenous Dirichlet data $\mathbf{u}^D = 0$ it holds

$$\|\mathbf{u}\|_{H^2(\Omega)^3 \cap H_0^1(\Omega)^3} \leqslant c \|\mathbf{f}\|_{L^2(\Omega)^3}.$$

Regularity of solutions is usually restricted at points, where Neumann and Dirichlet parts of the boundary come together. Here, we usually have singularities in the gradient of the solution and the stress tensor.

The incompressible Navier-Lamé equations

For incompressible linear materials with $\nu = \frac{1}{2}$, the stress tensor is reduced to

$$\boldsymbol{\sigma} = \mu(\nabla \mathbf{u} + \nabla \mathbf{u}^T),$$

as $\mathrm{tr}(\boldsymbol{\epsilon}) = \mathrm{div}\mathbf{u} = 0$. The material is no longer able to react on purely isotropic stresses. To formulate the incompressible Navier-Lamé equations, we consider a minimization problem in the space of divergence free functions

$$\mathbf{u} \in V_0 : \quad E(\mathbf{u}) \leqslant E(\mathbf{v}) = \frac{1}{2}a(\mathbf{v},\mathbf{v}) - l(\mathbf{v}) \quad \forall \mathbf{v} \in V_0,$$

where $a(\mathbf{v},\mathbf{v}) = (\boldsymbol{\sigma}, \nabla\mathbf{v})$, $l(\mathbf{v}) = (\mathbf{f},\mathbf{v}) + \langle u^{\sigma}, \mathbf{v}\rangle_{\Gamma^N}$ and where V_0 is the space of weakly divergence free functions

$$V_0 = \{\phi \in H_0^1(\Omega;\Gamma^D)^3,\ (\mathrm{div}\phi, \xi) = 0\ \forall \xi \in L^2(\Omega)\}. \tag{1.37}$$

The Hilbert space V_0 is a closed subspace of $H_0^1(\Omega;\Gamma^D)^3$, such that the existence of a unique solution follows as shown in Lemma 1.28. To derive a variational formulation, we use the Euler-Lagrange approach for constraint minimization problems and define the Lagrange functional

$$\mathcal{L}(\mathbf{u},p) = \frac{1}{2}a(\mathbf{u},\mathbf{u}) - l(\mathbf{u}) - (p, \mathrm{div}\mathbf{u}),$$

with a Lagrange multiplier $p \in L^2(\Omega)$. A possible solution is given as stationary point of $\mathcal{L}(\mathbf{u},p)$:

$$\begin{aligned} d_{\mathbf{u}}\mathcal{L}(\mathbf{u},p)(\phi) &= a(\mathbf{u},\phi) - l(\phi) - (p,\mathrm{div}\phi) &&\overset{!}{=} 0 \quad \forall \phi \in H_0^1(\Omega;\Gamma^D)^3 \\ d_p\mathcal{L}(\mathbf{u},p)(\xi) &= -(\xi,\mathrm{div}\mathbf{u}) &&\overset{!}{=} 0 \quad \forall \xi \in L^2(\Omega). \end{aligned}$$

We include the Lagrange multiplier into the stress tensor and define

$$\boldsymbol{\sigma}_I(\mathbf{u},p) = -pI + \mu(\nabla\mathbf{u} + \nabla\mathbf{u}^T),$$

where we identify $p \in L^2(\Omega)$ with a *pressure function*. This identification is reasonable, as $-pI$ acts as isotropic stress in all directions. The problem is now to find $\{\mathbf{u},p\} \in H_0^1(\Omega;\Gamma^D)^3 \times L^2(\Omega)$ such that

$$\big(\mu(\nabla\mathbf{u} + \nabla\mathbf{u}^T) - pI, \boldsymbol{\epsilon}(\phi)\big) + (\mathrm{div}\mathbf{u}, \xi) = (\mathbf{f},\phi) + \langle u^{\boldsymbol{\sigma}}, \phi\rangle_{\Gamma^N} \tag{1.38}$$

for all $\phi \in H_0^1(\Omega;\Gamma^D)^3$ and $\xi \in L^2(\Omega)$.

The incompressible Navier-Lamé equations, as a minimization problem with side condition is a *saddle-point system*. Existence and uniqueness theory cannot be based on ellipticity (in p). Instead, we split the proof for the existence of a well defined solution in two parts. We start by finding a suitable deformation field. Therefore, we restrict the space of admissible functions to those that already fulfill the divergence condition in the space V_0, see (1.37). Then, it holds

Lemma 1.31 (Incompressible Navier-Lamé - Existence of unique solutions (displacement))**.** Let $\mathbf{f} \in L^2(\Omega)^3$, $\bar{\mathbf{u}}^D \in H^1(\Omega)^3$ be an extension of the Dirichlet data into the domain and $\mathbf{u}^{\boldsymbol{\sigma}} \in H^1(\Gamma^N)^3$. There exists a unique solution $u \in \bar{u}^D + H_0^1(\Omega;\Gamma^D)^d$ to the variational problem

$$(2\mu\epsilon(\mathbf{u}), \epsilon(\phi)) = (\mathbf{f},\phi) + \langle \mathbf{u}^{\boldsymbol{\sigma}}, \phi\rangle_{\Gamma^N} \quad \forall \phi \in H_0^1(\Omega;\Gamma^D)^3.$$

For this solution it holds

$$\|\mathbf{u}\|_{H^1(\Omega)} \leqslant c \left(\|\mathbf{f}\|_{L^2(\Omega)} + \|\mathbf{u}^D\|_{L^2(\Gamma^D)} + \|\mathbf{u}^\sigma\|_{H^1(\Gamma^N)} \right).$$

Finally, $\mathbf{u} \in V_0$ minimizes the energy function in the space V_0

$$E(\mathbf{u}) \leqslant E(\mathbf{v}) \quad \forall \mathbf{v} \in V_0.$$

Proof. The subspace $V_0 \subset H_0^1(\Omega; \Gamma_D)^3$ is a Hilbert-space. The variational formulation is V_0-elliptic and the existence of a unique solution as well as the a priori estimate follow in the same way as shown in Lemma 1.28. □

Next, given a deformation field $u \in V_0$ we find a corresponding pressure by analyzing the equation

$$p \in L^2(\Omega): \\ -(p, \nabla\phi) = (\mathbf{f}, \phi) + \langle \mathbf{u}^\sigma, \phi \rangle_{\Gamma^N} - (2\mu\boldsymbol{\epsilon}(\phi), \nabla\phi) \quad \forall \phi \in H_0^1(\Omega; \Gamma^D)^3.$$

Existence of solutions to this problem cannot be shown by simple variational arguments. Instead, we will define by

$$\langle \operatorname{grad} p, \phi \rangle := -(p, \nabla \cdot \phi) \quad \forall \phi \in H_0^1(\Omega; \Gamma^D)^3,$$

the weak gradient operator $-\operatorname{grad} = \operatorname{div}^* : L^2(\Omega) \to H^{-1}(\Omega)$ and show existence by proving surjectivity of $-\operatorname{grad}$ in appropriate function spaces. We postpone this discussion to Section 2, where we will come across the same pressure problem concerning the incompressible Stokes equations.

The non-stationary Navier-Lamé equations

The non-stationary system of Navier-Lamé equations as given in Definition 1.23 is a hyperbolic problem

$$\partial_{tt}\mathbf{u} - \operatorname{div}(\boldsymbol{\sigma}) = 0, \quad \mathbf{u}(0) = \mathbf{u}^0, \quad \partial_t\mathbf{u}(0) = \mathbf{v}^0.$$

For simplicity we will consider the case of homogenous Dirichlet data only and we will further assume that $\mathbf{f} = 0$. We multiply the differential equation by $\phi = \partial_t\mathbf{u}$ and integrate over the spatial domain to get

$$0 = (\partial_{tt}\mathbf{u}, \partial_t\mathbf{u}) + (\boldsymbol{\sigma}(\mathbf{u}), \boldsymbol{\epsilon}(\partial_t\mathbf{u})) = \frac{d}{dt} \left(\underbrace{\frac{1}{2}\|\partial_t\mathbf{u}\|^2 + \frac{1}{2}(\boldsymbol{\sigma}(\mathbf{u}), \boldsymbol{\epsilon})}_{=:E(t)} \right),$$

where by $E(t)$ we denote the energy of the system. This energy does not change over time (remember that we consider the homogenous problem only). Integration over the temporal domain $I = [0, T]$ yields the relation

$$E(t) = E(0) \quad t \geqslant 0, \quad E(0) = \frac{1}{2}\|\mathbf{v}^0\|^2 + \frac{1}{2}\left(\boldsymbol{\sigma}(\mathbf{u}^0), \boldsymbol{\epsilon}(\mathbf{u}^0)\right),$$

with the initial velocity $\mathbf{v}^0 = \partial_t \mathbf{u}^0$. Hence a solution must by unique and it is bounded by the initial data.

The conservation of energy $d_t E(t) = 0$ shows the close relation to the wave equation. Existence of solutions to this simple (linear, symmetric and positive) problem can be shown by the Fourier approach. The operator

$$\langle \mathcal{L}\mathbf{u}, \mathbf{v} \rangle := \Big(2\mu \boldsymbol{\epsilon}(\mathbf{u}) + \lambda \operatorname{tr}\big(\boldsymbol{\epsilon}(\mathbf{u})\big), \boldsymbol{\epsilon}(\mathbf{v}) \Big)$$

is symmetric, positive definite, selfadjoint and a bijection. Its inverse is bound and considered as operator $\mathcal{L}^{-1} : L^2(\Omega)^d \to L^2(\Omega)^d$ it is compact. Hence $\mathcal{L}$ has a spectrum of positive eigenvalues, with no finite accumulation point. Further, an orthonormal basis of eigenvectors exists. This allows to diagonalize the system of equations, such that it decomposes into a sequence of scalar initial values problems that have a solution that can be constructed by elementary principles. For the details on this construction, we refer to the literature [29].

A recent result on the regularity of the non-stationary Navier-Lamé problem with homogenous Dirichlet data is given by Mitrea and Monniaux [28]. They basically show that given sufficient regularity of the domain's boundary (Lipschitz), the solution of the non-stationary Navier-Lamé problem with zero initial data and zero Dirichlet data satisfies $\mathbf{u} \in H^1(I; L^2(\Omega)^3)$ for every right hand side $\mathbf{f} \in L^2(I; L^2(\Omega)^3)$.

In the upcoming chapters, we will see that the coupling of the solid equation to the fluid equations brings along further challenges for the analysis of the partial differential equations. The *kinematic coupling condition*, see Section **??** will ask for continuity of solid- and fluid-velocities on a common interface $\mathcal{I}(t) = \partial\mathcal{S}(t) \cap \partial\mathcal{F}(t)$

$$\mathbf{v}_f = \mathbf{v}_s \text{ on } \mathcal{I}(t).$$

In the case of stationary problems, this kinematic coupling condition is just a usual no-slip boundary condition $\mathbf{v}_f = 0$ for the fluid's velocity. For fully non-stationary problems, a real coupling between the two velocities is introduced. The solution of the Navier-Stokes equations is well defined for velocities with traces in

$$\mathbf{v}_f\Big|_{\partial\mathcal{F}} \in H^{\frac{1}{2}}(\partial\mathcal{F}),$$

which – as seen from the solid problem – will require

$$\mathbf{v}_f\Big|_{\mathcal{I}} = \mathbf{v}_s\Big|_{\mathcal{I}} \quad \Rightarrow \quad \mathbf{v}_s \in H^1(\mathcal{S}).$$

However, the previous analysis only gives

$$\mathbf{v}_s = \partial_t \mathbf{u}_s \in L^2(I; L^2(\Omega)^3).$$

This is not sufficient to define a $H^{1/2}$-trace on $\mathcal{I}$. This problem has two possible solutions. First – and this will be our usual procedure – we can simply assume additional a priori knowledge on the regularity of $\mathbf{u}_s$ and therefore $\mathbf{v}_s$. This can be guaranteed for small and

regular problem data, if the boundaries of the coupled problem have very high regularity. Coutand and Shkoller [10] show the existence of solutions for the coupling of elastic solids with the Navier-Stokes equations, if the solid with boundary of class H^4 is completely embedded in a fluid-domain with boundary of class H^3, given sufficient regularity of the right hand side and the boundary data, see [10]. A second approach to enforce sufficient regularity it to add damping terms to the solid equation. Gazzola and Squassina show the following result, see [14].

Theorem 1.32 (Damped wave equation)**.** Let $\Omega \subset \mathbb{R}^d$ be a Lipschitz domain. The strongly damped wave equation

$$\partial_{tt} u - \Delta u - \omega \Delta \partial_t u + \mu \partial_t u = 0 \text{ in } [0,T] \times \Omega,$$

with initial values

$$u(0,\cdot) = u_0 \in H^1(\Omega), \quad \partial_t u(0,\cdot) = u_1 \in L^2(\Omega),$$

and homogenous Dirichlet values on $\partial\Omega$ and the damping parameters

$$\omega > 0, \quad \mu > -\omega\lambda_1,$$

where λ_1 is the first eigenvalue of $-\Delta$ has a unique solution satisfying

$$u \in L^\infty([0,T], H_0^1(\Omega)) \cap W^{1,\infty}([0,T], L^2(\Omega)), \quad \partial_t u \in L^2([0,T], H_0^1(\Omega)).$$

For the proof, see Gazolla and Squassina [14].

By adding strong damping terms, we are able to assure sufficient regularity to realize the kinematic coupling condition between solid problem and fluid problem.

1.3.2 Theory of nonlinear hyper-elastic material

Tackling the existence and uniqueness problem of the full elastic structure equation (using the St. Venant Kirchhoff material law) is complicated by the nonlinearity of the problem. Here, we will not give details on the complex proofs, but will simply cite some important results. A good overview on the theory of nonlinear elastic materials is given in the textbook of Ciarlet [7].

All approaches for the nonlinear problem will at some time use a linearization of the problem and will consult the theory that has been derived for the linear Navier-Lamé problem. Further, most approaches use variational techniques, such that the starting point for every analysis is the following weak formulation of the problem:

Lemma 1.33 (Weak formulation of the hyper-elastic structures)**.** Let $\bar{\mathbf{u}}^D \in H^1(\hat{S})^d$ be an extension of the Dirichlet data on Γ^D into the domain Ω. If the solution

$$\hat{\mathbf{u}}_f \in \bar{\mathbf{u}}_f^D + H_0^1(\hat{\Omega}; \Gamma^D)^d$$

of the variational formulation

$$(\hat{\mathbf{F}}\hat{\mathbf{\Sigma}}_s, \hat{\nabla}\hat{\phi})_{\hat{S}} = (\rho_s^0 \hat{\mathbf{f}}_s, \hat{\phi}), \quad \forall \hat{\phi} \in H_0^1(\hat{\Omega}; \Gamma^D)^d, \tag{1.39}$$

has sufficient regularity $\hat{\mathbf{u}} \in C^2(\hat{\Omega}) \cap C(\hat{\Omega} \cup \Gamma^D) \cap C^1(\hat{\Omega} \cup \Gamma^D)$, it is also a solution to the classical formulation of the elastic structure equations (1.33) with Dirichlet data on Γ_s^D.

Using the implicit function theorem, Ciarlet [7] proofs the following result for weak solutions of the elastic structure equation governed by the St. Venant Kirchhoff material:

Lemma 1.34 (Stationary St. Venant Kirchhoff material)**.** Let $\Omega \subset \mathbb{R}^3$ be a domain with C^2-boundary. Then, for every $p > 3$ there exists a constant α such that for every $\mathbf{f} \in L^p(\Omega)^d$ with $\|\mathbf{f}\|_{L^p} \leqslant \alpha$ there exists a unique solution $\mathbf{u} \in W^{2,p}(\Omega)$ to the stationary elastic structure equation governed by the St. Venant Kirchhoff material.

For the proof, we refer to the literature [7].

1.4 The fluid problem

In fluid-dynamics, we describe the flow of particles in the Eulerian framework. Looking at a fixed coordinate $x \in \mathbb{R}^d$ we observe a particle $\hat{x}(x, t)$ that at time t is in position x. The fate of a single particle is of no interest.

We will only consider incompressible fluids, i.e. a given moving volume $V(t)$ will not change its size under motion:

$$d_t|V(t)| = 0, \quad t \geqslant 0.$$

Applying Reynolds' Transport theorem, Theorem 1.8 to the scalar $\Phi \equiv 1$ yields:

$$d_t|V(t)| = d_t \int_{V(t)} 1 \, dx = \int_{V(t)} \operatorname{div} \mathbf{v} \, dx.$$

Hence as $V(t)$ can be chosen arbitrarily, we deduce the point-wise equation for the incompressibility of a fluid, see also Section 1.2.3:

$$\operatorname{div} \mathbf{v} = 0. \tag{1.40}$$

Using this condition, conservation of mass (1.14) reduces to a transport equation for the fluid's density:

$$\partial_t \rho_f + (\mathbf{v} \cdot \nabla)\rho_f = 0. \tag{1.41}$$

For further simplification, we will restrict all our considerations to homogenous fluids, where the density at initial time $t = 0$ is constant in the complete volume $\rho_f(x, 0) = \rho_f^0(x) \equiv \rho_f$. Given (1.41) it hereby follows that the density is homogenous at all times $t \geqslant 0$ and conservation of mass is reduced to the divergence condition $\operatorname{div} \mathbf{v} = 0$.

To close the system of equations for incompressible fluids we must introduce material laws that model the dependency of the stress tensor $\boldsymbol{\sigma}_f$ on velocity and pressure. We are considering *Navier-Stokes* fluids only that linearly depend on the strain rate following Hooke's law

$$\boldsymbol{\sigma} = 2\mu_f \dot{\boldsymbol{\epsilon}} + \lambda \operatorname{tr}(\dot{\boldsymbol{\epsilon}}) I.$$

As for an incompressible fluid it holds $\operatorname{div} \mathbf{v} = \operatorname{tr}(\dot{\boldsymbol{\epsilon}}) = 0$, the stress tensor simplifies to

$$\boldsymbol{\sigma} = -pI + \mu_f(\nabla \mathbf{v} + \nabla \mathbf{v}^T), \tag{1.42}$$

where again by p we denote the undetermined pressure that will act as Lagrange multiplier to ensure the divergence condition $\operatorname{div} \mathbf{v} = 0$. By $\mu_f = \rho_f \nu_f$ we denote the dynamic viscosity of the fluid and by ν_f its kinematic viscosity. The complete set of the Navier-Stokes equations is given by

$$\rho_f(\partial_t \mathbf{v} + (\mathbf{v} \cdot \nabla)\mathbf{v}) - \operatorname{div} \boldsymbol{\sigma} = \rho_f \mathbf{f}, \quad \operatorname{div} \mathbf{v} = 0,$$

or, using the material law for a Navier-Stokes fluid

$$\rho_f \left(\partial_t \mathbf{v} + (\mathbf{v} \cdot \nabla)\mathbf{v}\right) + \nabla p - \rho_f \nu_f \operatorname{div}(\nabla \mathbf{v} + \nabla \mathbf{v}^T) = \rho_f \mathbf{f}, \quad \operatorname{div} \mathbf{v} = 0. \tag{1.43}$$

Remark 1.35 (Symmetry of the stress-tensor)**.** For an incompressible fluid, the stress-tensor allows for a further simplification. It holds:

$$\left[\operatorname{div}\left(\nabla \mathbf{v} + \nabla \mathbf{v}^T\right)\right]_i = \sum_j \partial_j \left(\partial_j \mathbf{v}_i + \partial_i \mathbf{v}_j\right) = \Delta \mathbf{v}_i + \partial_i \underbrace{\operatorname{div} \mathbf{v}}_{=0} \text{ for } i = 1, 2, 3,$$

and Equation (1.43) is equivalent to the reduced formulation

$$\rho_f \left(\partial_t \mathbf{v}_f + (\mathbf{v} \cdot \nabla)\mathbf{v}\right) - \rho_f \nu_f \Delta \mathbf{v} + \nabla p = \rho_f \mathbf{f}, \quad \operatorname{div} \mathbf{v} = 0.$$

Usually, this simplified set of equations is considered as the Navier-Stokes equations. However, while both equations yield the same solution $(\mathbf{v}, p)$, the value of boundary stresses might altered, if the reduced tensor $\tilde{\boldsymbol{\sigma}}_f = \mu_f \nabla \mathbf{v} - pI$ is considered

$$\tilde{\boldsymbol{\sigma}}_f \mathbf{n} = \boldsymbol{\sigma}_f \mathbf{n} + \rho_f \nu_f \nabla \mathbf{v}^T \mathbf{n}.$$

We consider the two dimensional case and a straight boundary with $\mathbf{n} = (0, 1)^T$

$$\tilde{\boldsymbol{\sigma}}_f \mathbf{n} = \boldsymbol{\sigma}_f \mathbf{n} + \rho_f \nu_f \begin{pmatrix} \partial_x \mathbf{v}_1 & \partial_x \mathbf{v}_2 \\ \partial_y \mathbf{v}_1 & \partial_y \mathbf{v}_2 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \boldsymbol{\sigma}_f \mathbf{n} + \rho_f \nu_f \begin{pmatrix} \partial_x \mathbf{v}_2 \\ \partial_y \mathbf{v}_2 \end{pmatrix}$$

The two boundary stresses are different, if the normal component $\mathbf{v}_2$ of the velocity differs from zero. Usually, this will not happen on a rigid wall of the domain. The fluid will not enter or leave an obstacle. If the wall however is moving (e.g. in the case of fluid-structure interaction problems) it holds $\mathbf{v}_2 \neq 0$ and the stresses will differ. We refer to the literature [31]. △

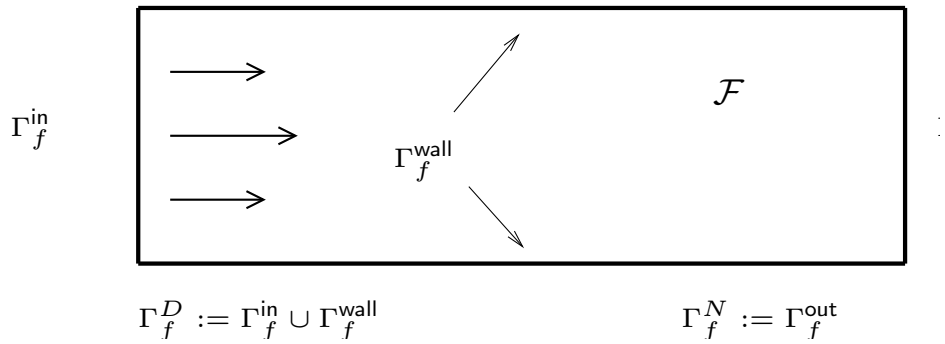


Figure 1.7: Typical configuration of a flow problem with Dirichlet inflow boundary Γ_f^{in} and Dirichlet no-slip boundary on the walls Γ_f^{wall} as well as an outflow boundary Γ_f^{out} of Neumann type.

1.4.1 Boundary and initial conditions

The system of equations is completed by adequate boundary and initial conditions. Let $\mathcal{F} \subset \mathbb{R}^d$ be the fluid-domain. At time $t = 0$ we prescribe an initial condition for the velocity

$$\mathbf{v}(x, 0) = \mathbf{v}^0(x) \quad x \in \mathcal{F}.$$

As the density is constant $\rho_f(x, t) \equiv \rho_f$ for all times (and homogenous in the domain), we do not need an initial condition here, but simply consider $\rho_f \in \mathbb{R}$ as a problem parameter. The boundary $\partial\mathcal{F}$ is split into a Dirichlet part Γ_f^D and into a Neumann part Γ_f^N. On Γ_f^D we prescribe Dirichlet conditions for the velocity

$$\mathbf{v}(x, t) = \mathbf{v}^D(x, t) \quad \text{on } \Gamma_f^D \times [0, T].$$

In the case $\mathbf{v}^D = 0$, we denote this condition as the *no-slip condition*. Physical observation tells us that viscosity will cause the fluid to stick to the boundary. This condition holds for the flow of water over elastic material (at usual velocities). The importance of viscous effects is lessened at high velocities, when e.g. considering the aerodynamical flow of air around a plane. Here, one often refers to the *slip condition* that only prescribes the flow in normal direction

$$\mathbf{n} \cdot \mathbf{v}(x, t) = 0 \quad \text{on } \Gamma_f^D \times [0, T].$$

The slip boundary condition prevents the flow from entering the boundary, it however allows for tangential flow. All examples considered in this work will be in the viscous regime where no-slip condition are usually well-placed. Boundaries with non homogenous Dirichlet data are often *inflow boundaries*.

Neumann conditions model situations, where we do not know the velocity profile at the boundary, but where assumptions on the boundary stress are given:

$$\boldsymbol{\sigma}_f(x, t)\mathbf{n}(x, t) = g^{\sigma}(x, t) \quad \text{on } \Gamma_f^N \times [0, T].$$

The typical application of Neumann conditions are *outflow boundaries*, where the profile of the flow is not known and a Dirichlet condition cannot be prescribed. See Figure 1.7 for a typical configuration of a flow problem with different boundary parts. We will come back to outflow boundary conditions in Section 1.4.2, as the exact form will depend on the material law and the Cauchy stress tensor $\boldsymbol{\sigma}_f$.

If only no-slip and outflow boundary conditions are taken into account, the complete set of incompressible flow equations on the (fixed) domain $\mathcal{F} \subset \mathbb{R}^d$ is given by

System 1.36 (Incompressible Navier-Stokes equations)**.** Velocity and pressure

$$\mathbf{v}(t) \in C^2(\mathcal{F}) \cap C(\mathcal{F} \cup \Gamma_f^D) \cap C^1(\mathcal{F} \cup \Gamma_f^N), \quad p(t) \in C^1(\mathcal{F}) \cap C(\mathcal{F} \cup \Gamma_f^N),$$

are given as solution of

$$\begin{aligned} \operatorname{div} \mathbf{v} = 0, \quad \rho_f \left(\partial_t \mathbf{v} + (\mathbf{v} \cdot \nabla)\mathbf{v}\right) &= \rho_f \mathbf{f} + \operatorname{div} \boldsymbol{\sigma}_f && \text{on } \mathcal{F} \times [0, T], \\ \mathbf{v}(\cdot, 0) &= \mathbf{v}^0(\cdot) && \text{on } \mathcal{F}, \\ \mathbf{v} &= \mathbf{v}^D && \text{on } \Gamma_f^D \times [0, T], \\ \boldsymbol{\sigma}_f \mathbf{n} &= g^\sigma && \text{on } \Gamma_f^N \times [0, T]. \end{aligned} \tag{1.44}$$

If boundary data $\mathbf{v}^D$ and g^σ as well as volume force $\mathbf{f}$ do not explicitly depend on time, the flow configurations can tend to a stationary limit, where it holds $\partial_t \mathbf{v} = 0$. Stationary in the context of fluid dynamics stands for a flow that at all times looks the same way, it does not imply that the fluid is at rest, which would mean $\mathbf{v} = 0$. If we know that the flow will reach a stationary limit, we can immediately consider the set of stationary equations, given as a boundary value problem.

System 1.37 (Stationary incompressible Navier-Stokes equations)**.** Velocity and pressure

$$\mathbf{v} \in C^2(\mathcal{F}) \cap C(\mathcal{F} \cup \Gamma_f^D) \cap C^1(\mathcal{F} \cup \Gamma_f^N), \quad p \in C^1(\mathcal{F}) \cap C(\mathcal{F} \cup \Gamma_f^N),$$

are given as solution of

$$\begin{aligned} \operatorname{div} \mathbf{v} = 0, \quad \rho_f (\mathbf{v} \cdot \nabla)\mathbf{v} &= \rho_f \mathbf{f} + \operatorname{div} \boldsymbol{\sigma}_f && \text{on } \mathcal{F}, \\ \mathbf{v} &= \mathbf{v}^D && \text{on } \Gamma_f^D, \\ \boldsymbol{\sigma}_f \mathbf{n} &= g^\sigma && \text{on } \Gamma_f^N. \end{aligned} \tag{1.45}$$

Not all autonomous flow problems have a stationary limit. This stems from the nonlinearity of the Navier-Stokes equations and whether a flow is stationary or instationary will depend on the problem data like density, viscosity, right hand side f and inflow velocity $\mathbf{v}^D$.

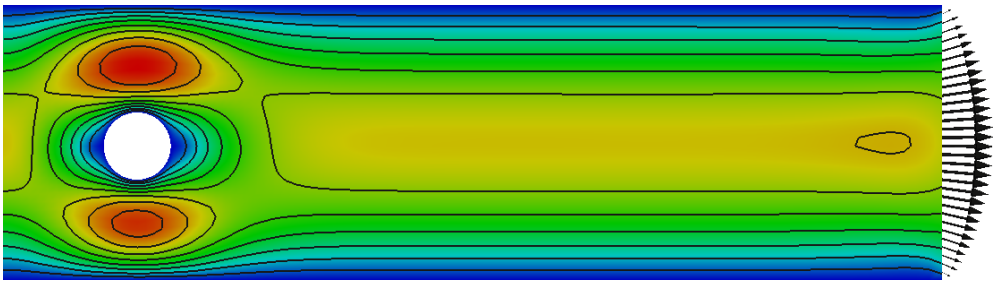


Figure 1.8: Channel flow with natural outflow condition $\boldsymbol{\sigma}_f \mathbf{n} = 0$. The velocity field gets deflected and does not follow the Poiseuille flow.

1.4.2 The "do-nothing" outflow condition

Many problem configurations feature boundaries, where the flow has mainly an outflow-character. We will call this boundary Γ_f^{out}. Here, the solution is not known a priori and cannot be specified in terms of a Dirichlet condition. Any boundary condition that is enforced, will be a model for the flow at the outflow boundary. Hence a common practice is to not describe a condition at all, but simply use the "natural" boundary condition, that arises from integration by parts. We consider the stationary Stokes equations:

$$(\boldsymbol{\sigma}_f, \nabla\phi)_{\mathcal{F}} = -(\mathrm{div}\boldsymbol{\sigma}_f, \phi)_{\mathcal{F}} + \langle \boldsymbol{\sigma}_f \mathbf{n}, \phi \rangle_{\Gamma_f^{out}},$$

from where we can deduce the "outflow-condition"

$$\boldsymbol{\sigma}_f \mathbf{n} = 0 \text{ on } \Gamma_f^{out}.$$

In Figure 1.8, we show a solution to a "channel-flow" problem using this natural outflow-condition. The domain is a channel with length L and height H:

$$\mathcal{F} = (0, L) \times (0, H),$$

on the left boundary Γ_f^{in} we impose a Dirichlet inflow profile

$$\mathbf{v} = \mathbf{v}^D = \frac{4\bar{\mathbf{v}}}{H^2} \begin{pmatrix} y(H-y) \\ 0 \end{pmatrix} \text{ on } \Gamma_f^{in} = 0 \times (0, H), \tag{1.46}$$

where $\bar{\mathbf{v}}$ is the peak velocity. On the horizontal lines Γ_f^{wall} we impose homogenous Dirichlet conditions

$$\mathbf{v} = 0 \text{ on } \Gamma_f^{wall} = (0, L) \times 0 \cup (0, L) \times H.$$

The outflow boundary is given as

$$\Gamma_f^{out} = L \times (0, H).$$

In Figure 1.8 we see that the velocity vectors get deflected and swing out of line. Considering the outflow model $\boldsymbol{\sigma}_f \mathbf{n} = 0$, which simply states that no external stresses act, this behavior

can be interpreted as a duct that ends in an open space, such that the fluid can expand in all directions.

Often, computational domains are chosen simply as a restriction of a larger domain to an area where the interesting dynamics happen. Numerically, boundary lines often must be drawn to scale the problem down to a reasonable size. In such situations, a good outflow boundary should have as little influence on the solution as possible. Regarding Figure 1.8, the exact location of the outflow boundary should not change the flow pattern inside the domain. The natural condition does not satisfy this request.

One of the most simple analytical solutions to a channel problem is the *Poiseuille flow*. An extension of the inflow data (1.46) into the domain

$$\mathbf{v}(x,y) = \frac{4\bar{v}}{H^2}\begin{pmatrix} y(H-y) \\ 0 \end{pmatrix},$$

satisfies the Navier-Stokes equations in channels (without obstacle) together with the pressure field

$$p(x,y) = \frac{8\bar{v}}{H^2}x + c,$$

for every $c \in \mathbb{R}$. In channel-like situations as shown in Figure 1.8, an outflow condition should allow for Poiseuille flows without deterioration.

By a small modification of this outflow condition, we allow the Poiseuille flow to leave the domain without deflection. Using the reduced stress tensor introduced in Remark 1.35

$$\tilde{\sigma}_f = \rho_f \nu_f \nabla \mathbf{v} - pI,$$

it holds for the Poiseuille flow that

$$\tilde{\sigma}_f \mathbf{n} = (\mathbf{n}\cdot\nabla)\mathbf{v} - p\mathbf{n} = 0 \text{ on } \Gamma_f^{out}.$$

This condition is called the *do-nothing outflow condition*, as it has as little impact on the flow as possible (or as it is the natural boundary condition, that arises without doing anything, when using the reduced tensor), see [20]. In Figure 1.9, we show the flow around a cylinder using this do-nothing condition. Here, he streamlines leave the domain in a straight way. Compare Figure 1.8.

Remark 1.38 (Outflow conditions)**.** We must stress that the *do-nothing* outflow condition is not the better condition from a physical point of view. It is simply a model that allows for some standard flow situations like Poiseuille flow or Couette flow to reduce the sensitivity of the solution on the position of artificial boundaries. From a good outflow condition we expect that it has as little influence on the flow field as possible. If the outflow boundary is far away from a region of interest (e.g. from an obstacle) we expect that the flow close to the obstacle is not influenced by the position of the outflow boundary, if the outflow boundary condition does a good job. The *do-nothing* condition works excellent in several configurations. It does not only allow Poiseuille or Couette flows to leave the domain, it further allows

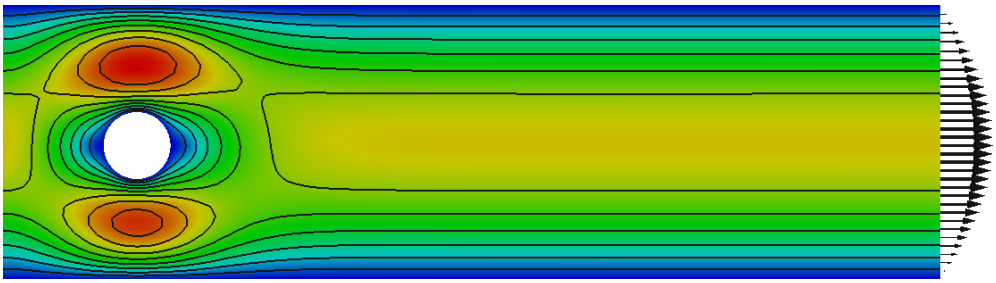


Figure 1.9: Channel flow with the *do-nothing outflow condition* $\rho_f \nu_f \nabla \mathbf{v}\mathbf{n} - pI = 0$ on Γ_f^{out}. The streamlines are not deflected on the right outflow boundary. Compare Figure 1.8.

vortices to leave the domain and has very small influence on these vortices, if the boundary is artificially cutting through them. However, many situations exist, where the analysis of outflow conditions is still not sufficiently developed: whenever the outflow boundary is not a single straight line normal to the main flow-direction, it will cause a deflection of the flow field. Further, if one considers more general material laws of non-Newtonian fluids, the *do-nothing* condition has an impact on the flow-field, see [42]. △

The *do-nothing* boundary condition brings along a further "hidden" boundary condition that normalizes the pressure. It can be shown [20] that on every straight outflow boundary-line segment $\Gamma_i \subset \partial\mathcal{F}$ that is enclosed by no-slip Dirichlet boundaries, it holds

$$\int_{\Gamma_i^{out}} p \, ds = 0,$$

on all outflow boundaries Γ_i^{out}, such that the average outflow pressure is zero. To show this relation, we consider a configuration like given in Figure 1.8. The boundary as the normal $\mathbf{n} = (1,0)^T$ such that is holds (for simplicity $\rho_f = \nu_f = 1$)

$$\nabla \mathbf{v}\mathbf{n} - p\mathbf{n} = \begin{pmatrix} \partial_x \mathbf{v}^1 - p \\ \partial_x \mathbf{v}^2 \end{pmatrix} = 0.$$

Next, we consider the intergral over the boundary Γ_f^{out} and assume, that is spreads in y-direction from 0 to H. For the first component we get

$$\int_0^H \partial_x \mathbf{v}^1(L,y) - p(L,y) \, dy = 0.$$

where L is the x-position of the outflow boundary. Using the divergence freeness of the flow $\partial_x \mathbf{v}^1 + \partial_y \mathbf{v}^2 = 0$ this is equivalent to

$$\int_0^H -\partial_y \mathbf{v}^2(L,y) - p(L,y) \, dy = -\Big(\mathbf{v}^2(L,H) - \mathbf{v}^2(L,0) - \int_0^H p(L,y) \, dy\Big) = 0.$$

If we now consider that homogenous Dirichlet conditions $\mathbf{v} = 0$ hold on the top and bottom wall the hidden condition is revealed

$$\int_{\Gamma_f^{out}} p ds = 0.$$

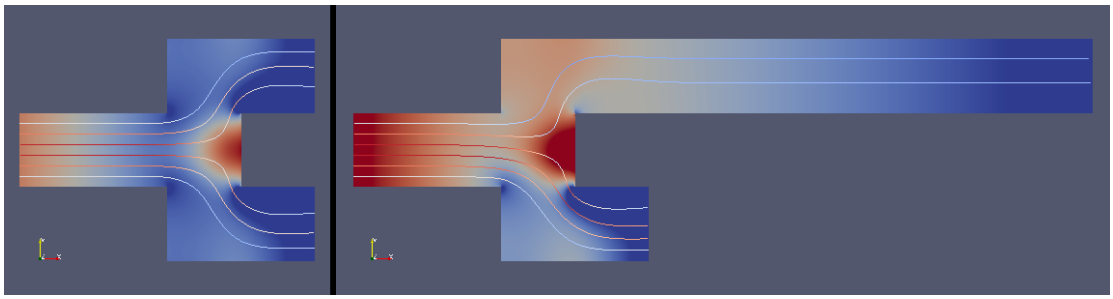


Figure 1.10: The *do-nothing* outflow condition for a domain with several outflow boundaries. The normalization of the pressure $\int_{\Gamma_i^{out}}$ on every outflow part leads to a deflection of the fluid.

This condition has two implications: first, whenever an outflow boundary of *do-nothing* type is given, no pressure-normalization has to be included in the trial spaces. Second, the *do-nothing* condition can be used to prescribe pressure drops on boundary segments in order to drive the flow:

$$\int_{\Gamma_i} \left(\rho_f \nu_f \mathbf{n} \cdot \nabla \mathbf{v} - p\mathbf{n}\right) ds = \int_{\Gamma_i} P_i \, ds, \quad i = 1, \ldots, N^{out}, \quad P_i \in \mathbb{R}.$$

This gets important, if the flow is driven by pressure differences and not by means of Dirichlet conditions. A frequently considered situation arises in hemodynamical simulations in which a flow in a part of the channel-system (i.e., the cardiovascular system) is investigated. This small part of the overall problem can be coupled by prescribing pressure values, e.g. taken from the pressure profile as measured from the heart-beat, see Figure 1.4.2.

1.4.3 The Reynolds number

The classical approach to fluid dynamics is the experiment: to determine the frictional forces of a ship we build a reduced model and test it in a flow channel. Transferring the results to the real scaled situation is not trivial. This is mainly due to the nonlinearity in the Navier-Stokes equations. (The linear Stokes equation will hardly every give realistic results). We will derive a *dimensionless* form of the Navier-Stokes equations.

The typical unit for the velocity $[v] = m/s$, for the density $[\rho_0] = kg/m^3$, for the pressure $[p] = kg/(m\,s^2)$ and for the viscosity $[\nu] = m^2/s$. To describe a flow configuration we introduce a characteristic length L and a characteristic speed V. An example for L could be the length of the ship, V could be the speed of the ship compared to the still ocean (Figure 1.11). Using these characteristic values we define the *dimensionless quantities*

$$x^* := \frac{1}{L}x, \quad v^* := \frac{1}{V}\mathbf{v}, \quad t^* := \frac{V}{L}t, \quad p^* := \frac{1}{\rho_0 V^2}p.$$

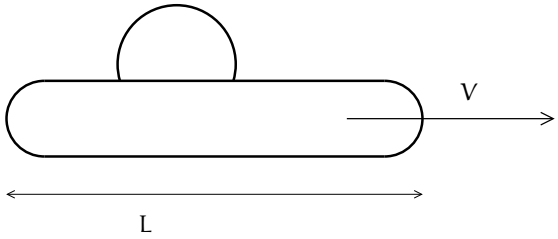


Figure 1.11: Definition of reference values for velocity V and length L. Further we need the model parameters density ρ and viscosity ν. For characterization of the flow configuration we derive the Reynolds number $R = LV/\nu$.

It holds

$$\frac{\partial \mathbf{v}^*}{\partial t^*} = \frac{L}{V^2}\partial_t \mathbf{v}, \quad \Delta^* \mathbf{v}^* = \frac{L^2}{V}\Delta \mathbf{v}, \quad (v^* \cdot \nabla^*)v^* = \frac{L}{V^2}(v \cdot \nabla)v, \quad \nabla^* p^* = \frac{L}{\rho_0 V^2}\nabla p.$$

We insert these values to the incompressible Navier-Stokes equations to get

$$\partial_{t^*}\mathbf{v}^* + (v^* \cdot \nabla^*)v^* = \frac{\nu}{LV}\Delta^* \mathbf{v}^* - \nabla^* p^* + \frac{L}{V^2} f.$$

By f we denote an outer volume force like the gravity force. On the left side of the equation we collect all inertia terms acting by the acceleration of the fluid. All quantities appear in dimensionless form. To describe the flow configuration we introduce two dimensionless stehen die Tr"aparameters, the *Reynolds number*

$$Re := \frac{LV}{\nu} = \frac{\rho_0 LV}{\mu},$$

tath indicates the relation between friction and inertia and the *Froude number*

$$Fr := \frac{V^2}{Lf},$$

indicating the relation between gravity and inertia. We will mostly deal with flow situations where the gravity is of lesser importance such that the Reynolds number will be the key parameter in our considerations. We call two different flow condfigurations *similar*, if the geometry of both domains is similar and if the two characteristic parameters Reynolds number and Froude number are agree. Then, the solution $\{\mathbf{v}^*, p^*\}$ and $\{\mathbf{v}, p\}$ are similar in the sense

$$\mathbf{v}(x,t) = V\mathbf{v}^*(x/L, tV/L), \quad p(x,t) = V^2\rho_0 p^*(x/L, tV/L).$$

Example 1.39 (Similar solutions)**.** An small tanker with a length of $L \approx 150\mathrm{m}$ and a speed of $V = 36\,\mathrm{km/h} = 10\,\mathrm{m/s}$ is to be investigated in a flow channel. The kinematic viscosituy of water is given as $\nu = 1.5 \cdot 10^{-6}\,\mathrm{m^2/s}$. Hereby we compute the Reynolds number and the Froude number as

$$Re = \frac{LV}{\nu} = \frac{150 \cdot 10}{1.5 \cdot 10^{-6}} = 10^9, \quad Fr = \frac{V^2}{Lf} = \frac{10^2}{150 \cdot 9.81} \approx \frac{1}{15}.$$

Assume, the dimension of our flow channel allows us to test a model of $6\,\mathrm{m}$, i.e. in a relation $L_{\text{model}} : L = 1 : 25$.

If we want to realize the same Froude number we must satisfy

$$\frac{V_{\text{model}}^2}{L_{\text{model}}} = \frac{V^2}{L}$$

therefore $V_{\text{model}} : V = 1 : 5$ which gives $V_{\text{model}} = 2\,\mathrm{m/s}$. If we also try to match the Reynolds number we can only vary the viscosity as last free parameter:

$$Re = 10^9 = \frac{6 \cdot 2}{\nu_{\text{model}}} \Rightarrow \nu_{\text{model}} \approx 10^{-8}\mathrm{m}^2/\mathrm{s}$$

Such a fluid is not available. Chloroform with $\nu \approx 10^{-7}\mathrm{m}^2/\mathrm{s}$ gets close. But just think of a large swimming pool of sufficient size filled with Chloroform.

If we assume that gravity effects are of little importance, we skip the matching of the Froude number and only tune the Reynolds number. Considering a flow channel with water ($\nu = 1.5 \cdot 10^{-6}\mathrm{m}^2/\mathrm{s}$) the model velocity must satisfy

$$Re = 10^9 = \frac{6 \cdot V_{\text{model}}}{1.5 \cdot 10^{-6}} \Rightarrow V_{\text{model}} = 250\mathrm{m/s} = 900\mathrm{km/h}.$$

This is no real improvement.

This example shows, that flow channel experiments with a relevance are difficult to realize. In the following we list kinematic viscosities and densities of different materials:

Material	Viscosity m^2/s	Density kg/m^3
Air	$1.7 \cdot 10^{-5}$	1.2
Water 5° C	$1.5 \cdot 10^{-6}$	1000
Water 20° C	$1.0 \cdot 10^{-6}$	998
Water 25° C	$0.9 \cdot 10^{-6}$	997
Glycerol 20° C	$0.95 \cdot 10^{-3}$	1261
Blood	10^{-5}	1060
Honey	10^{-2}	1400
Chloroform	10^{-7}	1500

The exact identification of characteristic length and speed allows some freedom. Reynolds and Froude number mostly serve to compare flow configurations with a very similar design, e.g. rescaling of models. We give some examples.

Example 1.40 (Reynolds number)**.** A car of length $L = 4\,\mathrm{m}$ drives at $V = 15\,\mathrm{m/s}$. In air, this yields the Reynolds number $Re \approx 5 \cdot 10^6$. A starting or landing plane of $L = 50\,\mathrm{m}$ flies at $V = 100\,\mathrm{m/s}$ giving $Re \approx 3 \cdot 10^8$. Fast planes at high velocity require the consideration of different models for air, including compressibility (change of density) and lesser effect of friction. A

large tanker $L = 300\,\mathrm{m}$ goes with $V = 10\,\mathrm{m/s}$ through water. This gives a Reynolds number of $Re = 2 \cdot 10^9$. A fish of $L = 0.2\,\mathrm{m}$ at $V = 3\,\mathrm{m/s}$ brings it to $Re = 4 \cdot 10^5$. Blood flow within the heart of $L = 0.02\,\mathrm{m}$ at $V = 0.1\,\mathrm{m/s}$ carries the Reynolds number $Re = 200$. It is however inaccurate to consider blood as a linear viscous fluid. Large particles have an impact on the rheology. Nonlinear, so called *non Newtonian* fluid models must be considered. Finally, honey drops from a spoon of size $L = 0.01\,\mathrm{m}$ with a speed of $V = 0.01\,\mathrm{m/s}$ and reaches the Reynolds number $Re = 10^{-2}$. Here, we can neglect the nonlinearity $\mathbf{v} \cdot \nabla \mathbf{v}$ and the Stokes equation is a reasonable model.

1.4.4 Model configurations

Some few flow configurations allow for an analytical solution. Usually this is not possible on general complex domains.

Shear flow (Couette-Flow)

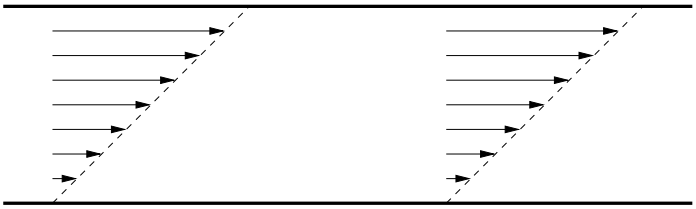


Figure 1.12: Velocity profile of the Couette flow.

We consider the flow in a gap between two infinite parallel planes $P_1 = (x, y, 0)$ and $P_2 = (x, y, L)$ with constant distance L. The upper plate P_2 moves relative to the lower plate within the $x - y$ plane and has the constant velocity $v = (v^x, v^y, 0)$. The velocity field

$$\mathbf{v}(x, y, z) = \frac{z}{L} \begin{pmatrix} v^x \\ v^y \\ 0 \end{pmatrix}.$$

is divergence free $\nabla \cdot \mathbf{v} = 0$ and together with the pressure $p = 0$ it satisfies the Navier-Stokes equations for all Reynolds numbers

$$-Re^{-1} \Delta \mathbf{v} + \mathbf{v} \cdot \nabla \mathbf{v} + \nabla p = 0, \quad \nabla \cdot \mathbf{v} = 0.$$

The Couette flow is the typical flow configuration where the pressure does not play a role.

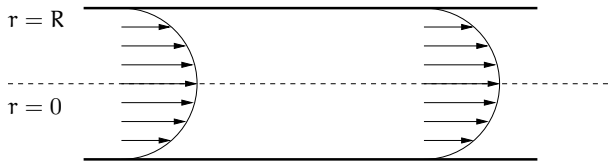


Figure 1.13: Velocity profile of the Poiseuille flow in a channel.

Channel flow (Poiseuille flow)

We consider a cylinder $\Omega \subset \mathbb{R}^3$ of infinite length. The x-axis is the middle line of the cylinder, the radius is given as $R > 0$. On the boundary of the cylinder we assume that the velocity is zero (no-slip condition). The velocity field

$$\mathbf{v}(x, y, z) = \begin{pmatrix} 1 - R^{-2}(y^2 + z^2) \\ 0 \\ 0 \end{pmatrix}$$

satisfies

$$\nabla \cdot \mathbf{v} = 0, \quad -\Delta \mathbf{v} = \begin{pmatrix} \frac{4}{r^2} \\ 0 \\ 0 \end{pmatrix}, \quad \mathbf{v} \cdot \nabla \mathbf{v} = \mathbf{0}.$$

Together with the pressure

$$p(x, y, z) = -4Re^{-1}r^{-2}x,$$

the pair $(\mathbf{v}, p)$ is a solution to the Navier-Stokes equations for all Reynolds numbers. The Poiseuille flow will not be found in experiments for high Reynolds numbers $Re > 5000$. Even though the Poiseuille flow is a mathematical solution, it is not stable and not physical. The Navier-Stokes-Gleichungen are nonlinear. For lage Reynolds numbers theory (and also the experiment) predict multiple solutions.

The Poiseuille flow is a model for the *Law of Hagen-Poiseuille*, which states that the flow rate (in blood vessels) goes with the forth power of the diameter. For a fixed pressure difference ΔP it holds in main flow direction (length L and radius R)

$$\frac{\Delta P}{L} \approx \Delta \mathbf{v} \approx \frac{1}{R^2},$$

therefore

$$\mathbf{v} \approx \frac{R^2 \Delta P}{L}.$$

Integration over an outflow plane Γ_{out} gives

$$\int_{\Gamma_{out}} \mathbf{n} \cdot \mathbf{v} \approx \frac{R^4 \Delta P}{L}.$$

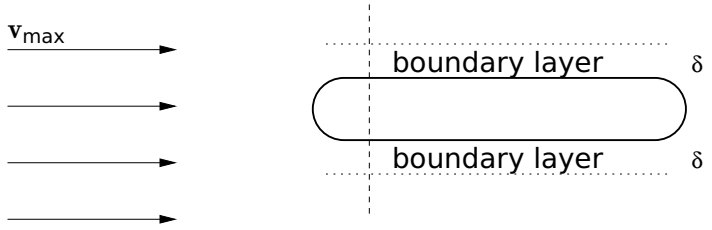


Figure 1.14: Boundary layer for the flow around an obstacle.

Boundary layers

A typical problem in fluid mechanics is to analyze the flow around a body of length L. At the boundary of such a body we assume that the fluid will stick due to viscous effects, i.e. $\mathbf{v} = 0$. In some distance to the obstacle the fluid will reach its main velocity $\mathbf{v} = \mathbf{v}_{max}$. The transition from this main flow to the no-slip boundary condition happens within a very small boundary layer. For moderate Reynolds numbers this boundary layer will feature streamlines parallel to the obstacle, see Figure 1.14. The width of the boundary layer depends on the Reynolds number. Close to the obstacle we observe higher x-velocities than y-velocities and the derivative in normal directions are dominant. We make the following simplifying assumptions

$$|v^x| \gg |v^y|, \quad |\partial_y \mathbf{v}| \gg |\partial_x \mathbf{v}| \tag{1.47}$$

and introduce new normalized variable within the boundary layer of width δ

$$\xi = \frac{x}{L}, \quad \eta = \frac{y}{\delta}.$$

Hereby we get

$$\partial_x \mathbf{v}^x = \frac{1}{L}\partial_\xi \mathbf{v}^x \approx \frac{v_{max}}{L}, \quad , \partial_y \mathbf{v}^x = \frac{1}{\delta}\partial_\eta \mathbf{v}^x \approx \frac{v_{max}}{\delta},$$

Conservation of mass $\partial_x \mathbf{v}^x + \partial_y \mathbf{v}^y = 0$ yields

$$\partial_y \mathbf{v}^y \approx \frac{v_{max}}{L},$$

and together with $\mathbf{v}(x, 0) = 0$ on the surface of the onstacle we get

$$\partial_y \mathbf{v}^y = -\partial_x \mathbf{v}^x \Rightarrow v^y(y) = -\int_0^y \partial_x \mathbf{v}^x dy \approx \frac{v_{max}}{L}\delta.$$

In terms of the derivatives this is

$$\partial_x \mathbf{v}^y \approx \frac{v_{max}}{L^2}\delta.$$

We insert these relations to the Navier-Stokes equations

$$\underbrace{v^x \partial_x \mathbf{v}^x}_{\frac{v_{max}^2}{L}} + \underbrace{v^y \partial_y \mathbf{v}^x}_{\frac{v_{max}^2}{L}} - \nu(\underbrace{\partial_x^2 \mathbf{v}^x}_{\frac{v_{max}}{L^2}} + \underbrace{\partial_y^2 \mathbf{v}^x}_{\frac{v_{max}}{\delta^2}}) + \frac{1}{\rho}\partial_x p = 0,$$

and would get a corresponding equation for the second component of the velocity. As $L \gg \delta$ we neglect the small term $\partial_x^2 v^x$. To satisfy the balance of forces between acceleration and friction the necessary condition reads

$$\frac{v_{max}^2}{L} \approx \nu \frac{v_{max}}{\delta^2} \Rightarrow \delta^2 \approx \frac{\nu L}{v_{max}} = \frac{L^2}{Re}$$

Therefore the width of the boundary layer must satisfy

$$\delta \approx \frac{L}{\sqrt{Re}}.$$

If we plan to numerically simulate a flow at Reynolds number $Re = 10^8$, the width of the boundary layer is as small as $\delta = 10^{-4}$. To resolve such a boundary layer by a discretization, very fine meshes are required. Fully resolved three dimensional simulations at high Reynolds numbers are not possible with standard techniques.

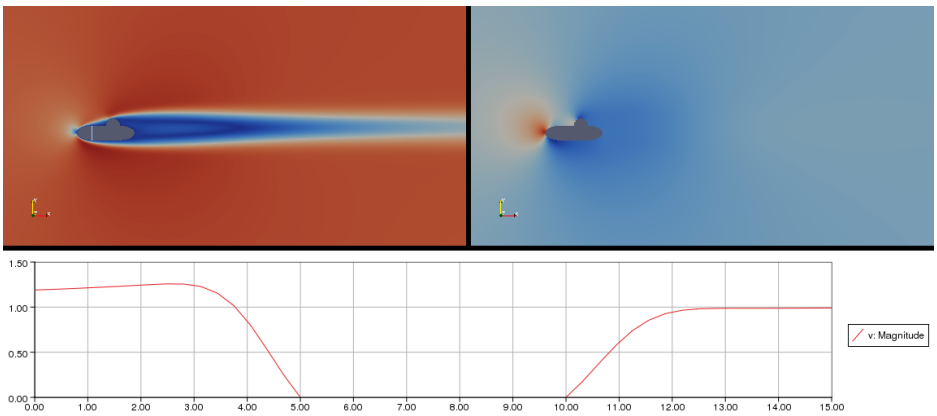


Figure 1.15: Stationary flow at Reynolds number $Re = 100$.

In figure 1.15 we show the flow around an obstacle at Reynolds number $Re = 100$. On the left side we show the velocity profile, right we depict the pressure. The velocity goes down from one (red) to zero (blue) in a small layer. In the bottom part we show the speed along a line that goes through the obstacle. The width of the boundary layer is about 1. This flow configuration at Reynolds number $Re = 100$ is stationary, velocity and pressure profile do not change in time. In front of the obstacle (left) we get high pressures, behind the obstacle the pressure is low. Here, a backflow is possible (called the wake, *Totwasser*).

Figure 1.16 shows the same configuration at increased Reybolds number $Re = 400$. Now, the result is a nonstationary flow. At the rear end of the obstacle, the boundary layer separates from the obstacle and vortices are created in the wake. These vortices are called *von Karman vortex street*. At moderate Reynolds numbers (such as here) the solution is periodic. We call stationary flows and nonstationary periodic flows *laminar*. Such flow patterns are characterized by stability: small perturbations are usually damped and do not completely change the flow pattern. Here the boundary layer is reduced to a width of about 0.5.

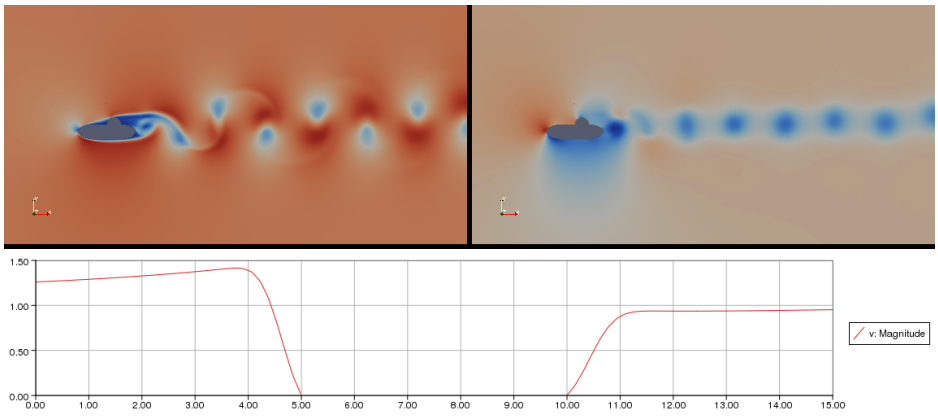


Figure 1.16: Nonstationary flow around an obstacle at Reybolds number $\mathrm{Re} = 400$.

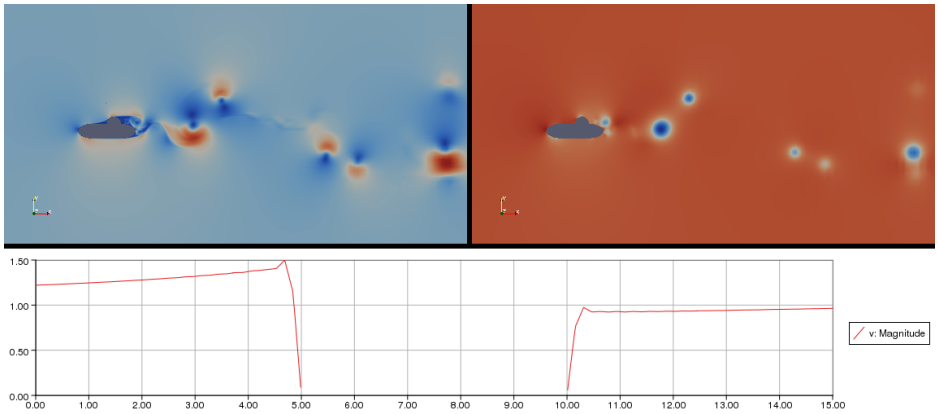


Figure 1.17: Nonstationary flow at $\mathrm{Re} = 12800$.

FInally we show in figure 1.17 the flow at Reynolds number $\mathrm{Re} = 12\,800$. This flow is in the critical regime and transition to a *turbulent* pattern is evolving. We cannot identify a clear periodic solution. Small changes (also small numerical perturbations) can change the complete flow field. Strong vortices appear in the domain. The size of the boundary layer is reduced to about 0.1.

This simple examples can be considered as the flow around a research submarine of $L = 20\,\mathrm{m}$ length. Considering the viscosity $\nu = 1.2 \cdot 10^{-6}$ the different Reynolds numbers still relate to very slow speeds

$$v = \frac{\nu \mathrm{Re}}{L}, \quad v_{100} \approx 2 \cdot 10^{-5}\mathrm{km/h}, \quad v_{400} \approx 8 \cdot 10^{-5}\mathrm{km/h}, \quad v_{12\,800} \approx 3 \cdot 6 \cdot 10^{-3}\frac{\mathrm{km}}{\mathrm{h}}.$$

In real situations, the flow around a submarine would require by far larger Reynolds numbers. Resolving the boundary layer would not be possible any more.

1.4.5 The stationary Navier-Stokes Equations

There are flow situations, where the velocity $\mathbf{v}$ and pressure p run into a *stationary limit* for $t \to \infty$. Then, velocity pressure will not change any more, i.e.

$$\partial_t \mathbf{v} = 0 \text{ and } \partial_t p = 0.$$

If we know that such a stationary limit exists and we are interested in simulating itm, we can directly consider the *stationary Navier-Stokes equations*.

$$\rho_f \mathbf{v} \cdot \nabla \mathbf{v} - \rho_f \nu \Delta \mathbf{v} + \nabla p = \rho \mathbf{f}, \quad \operatorname{div} \mathbf{v} = 0.$$

1.4.6 The linear Stokes Equations

In flow situations where friction effects are very large compared to acceleration terms, the Navier-Stokes equations can be simplified by neglecting the convective term $(\mathbf{v} \cdot \nabla)\mathbf{v}$. This case is given, if the Reynolds number tends to zero $Re \to 0$. If the right hand side of the equation as well as boundary data does not depend on time, the flow field will be stationary and we end up with the stationary Stokes equations

$$-\rho_f \nu_f \Delta \mathbf{v} + \nabla p = \rho_f \mathbf{f}, \quad \operatorname{div} \mathbf{v} = 0 \text{ in } \mathcal{F},$$

with the usual Dirichlet or Neumann boundary conditions on $\partial\mathcal{F}$. By renormalizing the pressure $\bar{p} = (\rho_f \nu_f)^{-1} p$ and the volume force $\bar{\mathbf{f}} = \nu_f^{-1} \mathbf{f}$ all physical parameters can be omitted and we derive the equations in non-dimensionalized form.

System 1.41 (Stokes Equations)**.** Velocity $\mathbf{v} \in C^2(\mathcal{F}) \cap C(\bar{\mathcal{F}})$ and pressure $p \in C^1(\mathcal{F})$ are given as solution of

$$-\Delta \mathbf{v} + \nabla \bar{p} = \bar{\mathbf{f}}, \quad \operatorname{div} \mathbf{v} = 0 \text{ in } \mathcal{F}. \tag{1.48}$$

Compared to the full incompressible Navier-Stokes equations, this equation is rather simple looking. As a saddle-point system it however still obtains one of the most important features of incompressible flows. While the physical relevance of the Stokes equations is very limited, it serves as entry-point to the mathematical analysis and the design of finite element discretizations for flow problems.

2 Theory of incompressible Flows

If there exists a unique solution $\{\mathbf{v}, p\}$ to the incompressible Navier-Stokes equations is still not known in all configuration. The stationary case is well understood, if we only consider Dirichlet boundary conditions. Here, a solution exists for small Reynolds numbers and it is unique, if the data is sufficiently small. When we consider general outflow conditions, we have no possibility to control the nonlinearity $(\mathbf{v}\cdot\nabla)\mathbf{v}$. In the instationary configuration there exists no proof for the existence of a unique solution under reasonable data assumptions. In three dimensions, the problem of proving the existence of a global smooth solution is considered open and one of the *Millenium Prize Problems*, see [6].

We start by deriving a weak formulation of the Navier-Stokes equations.

Lemma 2.1 (Weak formulation of the Navier-Stokes equations)**.** Let $\Omega \subset \mathbb{R}^d$ be a two $(d = 2)$ or three $(d = 3)$ dimensional domain with boundary $\partial\Omega = \Gamma^D \cup \Gamma^N$ that is split into Dirichlet part and Neumann (outflow) part. Let $\bar{\mathbf{v}}^D \in H^1(\Omega)^d$ be an extension of the Dirichlet data on Γ^D into the domain Ω. If the solution

$$\mathbf{v} \in \bar{\mathbf{v}}^D + \mathcal{V}, \quad \mathcal{V} := H_0^1(\Omega;\Gamma^D)^d, \quad p \in \mathcal{L}, \quad \mathcal{L} := L^2(\Omega),$$

of the variational formulation

$$\begin{aligned} (\rho(\partial_t\mathbf{v} + (\mathbf{v}\cdot\nabla)\mathbf{v}), \phi)_\Omega + (\rho\nu\nabla\mathbf{v}, \nabla\phi)_\Omega - (p, \nabla\cdot\phi)_\Omega = (\rho\mathbf{f}, \phi)_\Omega \qquad & \forall \phi \in \mathcal{V}, \\ (\operatorname{div}\mathbf{v}, \xi)_\Omega = 0 \qquad & \forall \xi \in \mathcal{L}, \end{aligned} \tag{2.1}$$

has sufficient regularity $v \in C^2(\Omega) \cap C(\Omega \cup \Gamma^D) \cap C^1(\Omega \cup \Gamma^{out})$ and $p \in C^1(\Omega)$, it also solves the classical formulation of the Navier-Stokes equations, Problem 1.36 with Dirichlet data on Γ^D and the *do-nothing* outflow condition on Γ^{out}.

Proof. This follows by integration by parts and with basic variational principles. The boundary term on Γ^{out} is required as we use the full symmetric stress-tensor such that the solution of the variational formulation fulfills the *do-nothing* condition, see Section 1.4.2. □

Remark 2.2 (Uniqueness of the pressure in Dirichlet problem)**.** If the configuration has Dirichlet boundaries all around the boundary $\Gamma^D = \partial\Omega$, the solution cannot be unique: let $\{\mathbf{v}, p\} \in \mathcal{V} \times \mathcal{L}$ be a solution. Then, it holds for $\{\mathbf{v}, p + c\}$ with $c \in \mathbb{R}$:

$$\begin{aligned} (\boldsymbol{\sigma},\nabla\phi)_\Omega &= \rho\nu(\nabla\mathbf{v}, \nabla\phi)_\Omega - (p + c, \nabla\cdot\phi)_\Omega \\ &= \rho\nu(\nabla\mathbf{v}, \nabla\phi)_\Omega - (p, \nabla\cdot\phi)_\Omega + (\underbrace{\nabla c}_{=0}, \phi)_\Omega - \langle cn, \underbrace{\phi}_{=0}\rangle_{\partial\Omega}. \end{aligned}$$

If $\Gamma^D = \partial\Omega$ the pressure can only be unique up to a constant. In this case, we normalize the pressure-space

$$\mathcal{L} = L^2(\Omega) \setminus \mathbb{R}.$$

If we consider open sets Ω that are not connected, we must filter the constants out of the pressure within every component. Here however, Ω will also be a domain, i.e. a connected open set. △

The Navier-Stokes equations brings along two characteristic difficulties for theoretical analysis and numerical discretization, the nonlinearity $(\mathbf{v} \cdot \nabla)\mathbf{v}$ and the side-condition of divergence freeness $\operatorname{div} \mathbf{v} = 0$. We will first focus on this second difficulty and consider the linear Stokes equations.

2.1 Existence and uniqueness of solutions to the stationary Stokes equations

In the following, we consider the stationary Stokes equations

$$\begin{aligned} \mathbf{v}, p \in \mathcal{V} \times L, \quad & \mathcal{V} := H_0^1(\Omega; \partial\Omega)^d, \quad \mathcal{L} := L^2(\Omega) \setminus \mathbb{R} : \\ & (\nabla \mathbf{v}, \nabla \phi)_\Omega - (p, \nabla \cdot \phi)_\Omega + (\nabla \cdot \mathbf{v}, \xi)_\Omega = (\mathbf{f}, \phi)_\Omega \quad \forall \{\phi, \xi\} \in \mathcal{V} \times \mathcal{L}. \end{aligned} \tag{2.2}$$

Here, we assume homogenous Dirichlet conditions on the complete boundary $\partial\Omega$ and further we consider the non-symmetric form of the stress tensor.

The Stokes equations are a *saddle point system*, the solution $\{\mathbf{v}, p\}$ cannot be written as a minimum of an optimization problem but as a saddle point of a constraint optimization problem.

Lemma 2.3 (Stokes as optimization problem)**.** Every solution $\{\mathbf{v}, p\} \in \mathcal{V} \times \mathcal{L}$ to the Stokes problem (2.2) is solution to the minimization problem

$$E(\mathbf{v}) \leqslant E(\phi) \quad \forall \phi \in \mathcal{V} := H_0^1(\Omega)^d, \quad E(\phi) := \frac{1}{2}\|\nabla \phi\|^2 - (\mathbf{f}, \phi)$$

under the constraint

$$(\operatorname{div} \mathbf{v}, \xi) = 0 \quad \forall \xi \in \mathcal{L} := L^2(\Omega) \setminus \mathbb{R}.$$

Proof. We formulate the constrained minimization problem with the Lagrangian

$$L(\mathbf{v}, p) := E(\mathbf{v}) - (p, \operatorname{div} \mathbf{v})$$

and aim for the stationary point

$$L'(\mathbf{v}, p)(\phi, \xi) = 0 \quad \forall \{\phi, \xi\} \in \mathcal{V} \times \mathcal{L}.$$

This (directional) derivative is defined as

$$L'(\mathbf{v},p)(\phi,\xi) = \frac{d}{ds}L(\mathbf{v}+s\phi,p)\Big|_{s=0} + \frac{d}{dt}L(\mathbf{v},p+t\xi)\Big|_{t=0}.$$

Here it gives

$$L'(\mathbf{v},p)(\phi,\xi) = (\nabla\mathbf{v},\nabla\phi) - (\mathbf{f},\phi) - (p,\operatorname{div}\phi) - (\xi,\operatorname{div}\mathbf{v}) \stackrel{!}{=} 0,$$

which is equivalent to the Stokes problem. □

Remark 2.4 (Saddle point problem)**.** The solution to the constraint minimization problem can be written as

$$L(\mathbf{v},\xi) \leqslant L(\mathbf{v},p) \leqslant L(\phi,p) \quad \forall\phi\in\mathcal{V},\ \xi\in\mathcal{L}.$$

Hereby we explain the label *saddle point problem*. △

By (2.2) we derive that every solution to the Stokes equation $\{\mathbf{v},p\}\in\mathcal{V}\times\mathcal{L}$ has a *divergence free* velocity $\mathbf{v}$. By divergence free "$\operatorname{div}\mathbf{v}=0$" we understand the variational condition

$$(\operatorname{div}\mathbf{v},\xi) = 0 \quad \forall\xi\in\mathcal{L}.$$

We introduce the space

$$\mathcal{V}_0 := \{\phi\in\mathcal{V} \,\big|\, (\operatorname{div}\phi,\xi) = 0\ \forall\xi\in\mathcal{L}\}.$$

Lemma 2.5 (Divergence-free functions)**.** The function space $\mathcal{V}_0$ is a closed subspace of $\mathcal{V}$. Functions $\mathbf{v}\in\mathcal{V}_0$ are called *solenoidal*.

Proof. Let $\mathbf{v}_n\in V_0$ be a Cauchy sequence of divergence free functions. Hence, it has a limit in $\mathcal{V}$

$$\mathbf{v}_n \to \mathbf{v}\in\mathcal{V}.$$

We show, that $\mathbf{v}$ is divergence free thus $\mathbf{v}\in\mathcal{V}_0$. It holds

$$(\operatorname{div}\mathbf{v},\xi) = (\operatorname{div}(\mathbf{v}-\mathbf{v}_n),\xi) + \underbrace{(\operatorname{div}\mathbf{v},\xi)}_{=0}.$$

We estimate with Cauchy Schwarz

$$\big|(\operatorname{div}\mathbf{v},\xi)\big| = \|\operatorname{div}(\mathbf{v}-\mathbf{v}_n)\|\,\|\xi\| \leqslant C\|\nabla(\mathbf{v}-\mathbf{v}_n)\|\,\|\xi\| \to 0 \quad (n\to\infty).$$

We used the estimate $\|\operatorname{div}\mathbf{v}\|\leqslant C\|\nabla\mathbf{v}\|$ that is shown in Lemma 2.6. □

Lemma 2.6. For $\mathbf{v}\in H^1(\Omega)^d$ it holds

$$\|\nabla\cdot\mathbf{v}\| \leqslant \sqrt{d}\|\nabla\mathbf{v}\|.$$

For $\mathbf{v}\in H_0^1(\Omega)^d$ it holds

$$\|\nabla\cdot\mathbf{v}\| \leqslant \|\nabla\mathbf{v}\|.$$

Proof. (i) It holds

$$\|\nabla \cdot \mathbf{v}\|^2 = \int_\Omega \left(\sum_{i=1}^{d} \partial_i \mathbf{v}_i \right)^2 dx = \sum_{i,j=1}^{d} \int_\Omega \partial_i \mathbf{v}_i \cdot \partial_j \mathbf{v}_j \, dx. \tag{2.3}$$

Using Young's inequality we get

$$\|\nabla \cdot \mathbf{v}\|^2 \leqslant \frac{1}{2} \sum_{i,j=1}^{d} \int_\Omega |\partial_i \mathbf{v}_i|^2 + |\partial_j \mathbf{v}_j|^2 \, dx \leqslant d \sum_{i=1}^{d} \int_\Omega |\partial_i \mathbf{v}_i|^2 \, dx \leqslant d \|\nabla \mathbf{v}\|^2.$$

(ii) Now, let $\mathbf{v} \in H_0^1(\Omega)^d$ with trace zero. We contintue with (2.3). Let $\phi_n \in C_0^\infty$ with $\phi_n \to \mathbf{v}_j$ in $H_0^1(\Omega)$ (this is possible, as the space C_0^∞ is dense in $H_0^1(\Omega)$). It holds with integration by parts

$$\int_\Omega \partial_i \mathbf{v}_i \cdot \partial_j \phi_n \, dx = \underbrace{\int_{\partial\Omega} \mathbf{n}_i \mathbf{v}_i \cdot \partial_j \phi_n \, do}_{=0} - \int_\Omega \mathbf{v}_i \cdot \partial_i \partial_j \phi_n \, dx.$$

We can change the order of differentiation to get

$$\int_\Omega \partial_i \mathbf{v}_i \cdot \partial_j \phi_n \, dx = - \underbrace{\int_{\partial\Omega} \mathbf{n}_j \mathbf{v}_i \cdot \partial_i \phi_n \, do}_{=0} + \int_\Omega \partial_j \mathbf{v}_i \cdot \partial_i \phi_n \, dx.$$

The limit $\phi_n \to \mathbf{v}_j$ transfers the result to the product $\partial_i \mathbf{v}_i \partial_j \mathbf{v}_j$ and use Young's inequality

$$\int_\Omega \partial_i \mathbf{v}_i \partial_j \mathbf{v}_j \, dx = \int_\Omega \partial_j \mathbf{v}_i \partial_i \mathbf{v}_j \, dx \leqslant \frac{1}{2} \int_\Omega |\partial_j \mathbf{v}_i|^2 \, dx + \frac{1}{2} \int_\Omega |\partial_i \mathbf{v}_j|^2 \, dx.$$

Summing over all i, j gives

$$\|\nabla \cdot \mathbf{v}\|^2 = \sum_{i,j=1}^{d} \int_\Omega \partial_i \mathbf{v}_i \cdot \partial_j \mathbf{v}_j \, dx \leqslant \frac{1}{2} \sum_{i,j=1}^{d} |\partial_j \mathbf{v}_i|^2 + |\partial_i \mathbf{v}_j|^2 \, dx = \|\nabla \mathbf{v}\|^2.$$

□

2.1.1 Existence and uniqueness of the velocity

A solution $\mathbf{v} \in \mathcal{V}$ to (2.2) will be weakly divergence free and thus in the space

$$\mathbf{v} \in \mathcal{V}_0 := \{\phi \in \mathcal{V}, \ (\operatorname{div} \phi, \xi)_\Omega = 0 \quad \forall \xi \in \mathcal{L}\} \subset \mathcal{V}.$$

By restricting the Stokes equations to this space, it remains to find

$$\mathbf{v} \in \mathcal{V}_0 : \quad (\nabla \mathbf{v}, \nabla \phi)_\Omega = (\mathbf{f}, \phi)_\Omega \quad \forall \phi \in \mathcal{V}_0. \tag{2.4}$$

Lemma 2.7 (Stokes velocity)**.** For every $\mathbf{f} \in H^{-1}(\Omega)^d$ there exists a unique velocity $\mathbf{v} \in \mathcal{V}_0 \subset \mathcal{V}$ as solution of the Stokes equations. Further, it holds

$$\|\nabla \mathbf{v}\| \leqslant \|\mathbf{f}\|_{-1}.$$

Proof. The space $\mathcal{V}_0 \subset \mathcal{V}$ is a vector space with the scalar product $(\nabla\cdot, \nabla\cdot)$. It is complete as shown in Lemma 2.5. Therefore, the existence of a unique solution follows by Riesz representation theorem (or Lax-Milgram). Further the estimate directly follows, see [33, 1]. □

2.1.2 Spectral theory for the Stokes operator

In the following, we will derive some basic properties of the Stokes operator in the space $\mathcal{V}_0$. We have shown that a velocity-solution $\mathbf{v} \in \mathcal{V}_0 \subset \mathcal{V}$ to the Stokes equations exists for every $\mathbf{f} \in H^{-1}(\Omega)$ and it holds

$$\|\nabla \mathbf{v}\| \leqslant \|\mathbf{f}\|_{H^{-1}(\Omega)}.$$

Given $\mathbf{f} \in L^2(\Omega)^d$ we can also employ the Cauchy-Schwarz estimate followed by Poincaré to get

$$\|\nabla \mathbf{v}\| \leqslant c_p \|\mathbf{f}\|.$$

By $\mathcal{J}_0$ we denote the space of weakly divergence free functions in L^2

$$\mathcal{J}_0 := \{\phi \in L^2(\Omega) \,\big|\, \nabla \cdot \phi = 0 \text{ and } \mathbf{n} \cdot \phi_{|\partial\Omega} = 0 \text{ weakly.}\} \tag{2.5}$$

Like $\mathcal{V}_0$ as closed subspace of $\mathcal{V}$ the space $\mathcal{J}_0$ is a closed subspace (with regard to the L^2-inner product and norm) of $L^2(\Omega)^d$. By

$$P_0 : L^2(\Omega)^d \to \mathcal{J}_0$$

we denote the orthogonal projection onto $\mathcal{J}_0$. This projection is called the *Helmholtz-Projection*. We cite the following theorem that is proven in [34].

Theorem 2.8. Let

$$\mathcal{J}_0 = \{\phi \in L^2(\Omega)^d \,\big|\, (\phi, \nabla\xi) = 0 \quad \forall \xi \in H^1(\Omega)\}$$
$$Z = \{\phi \in L^2(\Omega)^d \,\big|\, \exists \xi \in H^1(\Omega) \,:\, \phi = \nabla\xi\}.$$

Both spaces are closed subspaces of $L^2(\Omega)^d$ and it holds

$$L^2(\Omega)^d = \mathcal{J}_0 \oplus Z.$$

The orthogonal projection $P_0 : \mathcal{L}^2(\Omega) \to \mathcal{J}_0$ is bounded and it holds

$$\ker(P_0) = Z, \quad \operatorname{rg}(P_0) = \mathcal{J}_0.$$

By (2.5) an equivalent definition of $\mathcal{J}_0$ is given.

The following simple corollary holds

Corollary 2.9. Let $\mathbf{f} \in L^2(\Omega)^d$, $\mathbf{f}_0 := P_0\mathbf{f} \in \mathcal{J}_0$ the divergence free projection, $\mathbf{v} \in \mathcal{V}_0$ and $\mathbf{v}_0 \in \mathcal{V}_0$ be the corresponding velocity-solutions of the Stokes equations. It holds $\mathbf{v} = \mathbf{v}_0$.

Proof. As $\mathcal{J}_0 \subset \mathcal{V}_0$ it holds for all $\phi \in \mathcal{V}_0$

$$(\nabla\mathbf{v}, \nabla\phi) = (\mathbf{f}, \phi) = (P_0\mathbf{f}, \phi) = (\mathbf{f}_0, \phi) = (\nabla\mathbf{v}_0, \nabla\phi).$$

□

For $\mathbf{f} \in L^2(\Omega)^d$ we find $\mathbf{v} \in \mathcal{V}_0$ by

$$(\nabla\mathbf{v}, \nabla\phi) = (P_0\mathbf{f}, \phi) \quad \forall\phi \in \mathcal{V}_0.$$

As P_0 is bounded, the right hand side is a linear functional such that a unique solution $\mathbf{v}$ exists. We define the Stokes operator $S := -P_0\Delta$ as

$$S : D(S) \subset \mathcal{V}_0 \subset J_0 \to J_0.$$

The operator S is symmetric. Let $\mathbf{v} \in \mathcal{V}_0$ be the velocity solution to $\mathbf{f} \in \mathcal{J}_0$ and $\mathbf{w} \in \mathcal{V}_0$ the solution to $\mathbf{g} \in \mathcal{J}_0$:

$$(S\mathbf{v}, \mathbf{w}) = (\mathbf{f}, \mathbf{w}) = (\nabla\mathbf{v}, \nabla\mathbf{w}) = (\mathbf{v}, \mathbf{g}) = (\mathbf{v}, S\mathbf{w}).$$

Further it is positive definite

$$(S\mathbf{v}, \mathbf{v}) = (\mathbf{f}, \mathbf{v}) = \|\nabla\mathbf{v}\|^2 \geqslant c_p\|\mathbf{v}\|^2$$

and self-adjoint. The inverse Stokes operator $S^{-1} : \mathcal{J}_0 \to \mathcal{V}_0$ can be declared as a compact operator since the embedding $\mathcal{V}_0 \hookrightarrow \mathcal{J}_0$ is compact

$$S^{-1} : \mathcal{J}_0 \to \mathcal{J}_0.$$

Now it holds:

Theorem 2.10 (Spectral theorem for compact self-adjoint operators)**.** Let $T : H \to H$ a positive definite self-adjoint operator in the Hilbert space H. There exists an orthonormal system of Eigenvectors $e_1, e_2, \ldots$ and positive real Eigenvalues

$$\lambda_1 \geqslant \lambda_2 \geqslant \cdots > 0$$

that do not have a positive accumulation point. It holds

$$Tx = \sum_{k=1}^{\infty} \lambda_k \langle x, e_k\rangle e_k \quad \forall x \in H$$

and

$$\|T\| = \sup_k |\lambda_k|.$$

For a proof we refer to the literatur [43].

This theorem can be applied to the inverse Stokes operator. It shows that the inverse is bounded (which we already knew by the estimate $\|\nabla \mathbf{v}\| = \|\nabla(S^{-1}\mathbf{f})\| \leqslant c_p\|\mathbf{f}\|$ and that it allows for a representation in an L^2-orthonormal system of Eigenvectors. We will denote the orthonormal basis by $\mathbf{w}_1, \mathbf{w}_2, \cdots \in \mathcal{V}_0$. This basis is also a basis of $\mathcal{J}_0$ (as $\mathcal{V}_0$ is dense in $\mathcal{J}_0$). Hence, every $\mathbf{f} \in J_0$ can be represented as

$$\mathbf{f} = \sum_{k=1}^{\infty} \alpha_k \mathbf{w}_k, \quad \alpha_k := (\mathbf{f}, \mathbf{w}_k).$$

We will get back to these results when dealing with the nonlinear Navier-Stokes equations.

2.1.3 Existence and uniqueness of the pressure

The outline of this section closely follows the argumentation given by Schweizer [34]. In particular we will only give the complete proof for very limited class of domains Ω, namely quads $\Omega = (0, L)^2$ and cubes $\Omega = (0, L)^3$.

For every $\mathbf{f} \in H^{-1}(\Omega)$ we have a unique divergence free velocity $\mathbf{v} \in \mathcal{V}_0 \subset \mathcal{V}$. A corresponding pressure is determined as solution of the problem

$$p \in \mathcal{L}: \quad (p, \nabla \cdot \phi) = (\mathbf{f}, \phi) - (\nabla \mathbf{v}, \nabla \phi) \quad \forall \phi \in \mathcal{V}. \tag{2.6}$$

The corresponding bilinear form $b : \mathcal{L} \times \mathcal{V} \to \mathbb{R}$ defined by $b(p, \phi) := (p, \nabla \cdot \phi)$ is neither symmetric or coercive. Riesz representation theorem or generalizations like Lax-Milgram are not suitable. The right hand side of equation (2.6) defines a linear functional $j \in H^{-1}(\Omega)$ as for

$$j(\phi) := (\mathbf{f}, \phi) - (\nabla \mathbf{v}, \nabla \phi)$$

it holds with the a priori estimate for the velocity

$$\begin{aligned} \|j\|_{-1} := \sup_{\phi \in H_0^1(\Omega)^d} \frac{|j(\phi)|}{\|\nabla \phi\|} &= \sup_{\phi \in H_0^1(\Omega)^d} \frac{|(\mathbf{f}, \phi) - (\nabla \mathbf{v}, \nabla \phi)}{\|\nabla \phi\|} \\ &\leqslant \sup_{\phi \in H_0^1(\Omega)^d} \frac{(\|\mathbf{f}\|_{-1}\|\nabla \phi\| + \|\nabla \mathbf{v}\|\|\nabla \phi\|}{\|\nabla \phi\|} = \|\mathbf{f}\|_{-1} + \|\nabla \mathbf{v}\| \leqslant 2\|\mathbf{f}\|_{-1}. \end{aligned}$$

We reformulate this variational equation in operator notation as

$$-\operatorname{grad} p = j, \tag{2.7}$$

with

Definition 2.11 (Weak gradient)**.** The *weak gradient* is defined as operator

$$-\operatorname{grad} : \mathcal{L} \to H^{-1}$$

as

$$-\operatorname{grad}(p)(\phi) = \langle -\operatorname{grad} p, \phi \rangle = (p, \nabla \cdot \phi) \quad \forall \phi \in H_0^1(\Omega)^d.$$

Usually, we consider the gradient as a functional between spaces of higher regularity

$$-\nabla : H^1(\Omega) \to L^2(\Omega)^d.$$

Here, the solution to equation (2.7) is a L^2-function that does not even have the usual weak derivatives of the Sobolev space H^1.

To show existence of a unique pressure we must show that equation (2.7), i.e. $-\operatorname{grad} p = j$ has a unique solution $p \in L^2(\Omega)$ for all possible right hand sides j. This is equivalent to showing bijectivity of the weak gradient operator $-\operatorname{grad}$ in suitable spaces. Further, to show a desired bound on the solution $p \in \mathcal{L}$, i.e. $\|p\| \leqslant c\|\mathbf{f}\|_{-1}$ we require the property, that the weak gradient operator $-\operatorname{grad}$ is a bounded operator. We start by limiting the possible range of the weak gradient.

Lemma 2.12. Let $\mathbf{f} \in H^{-1}(\Omega)$ and $\mathbf{v} \in \mathcal{V}_0 \subset \mathcal{V}$ be the corresponding velocity of the Stokes operator. The functional

$$j(\phi) := (\mathbf{f}, \phi) - (\nabla \mathbf{v}, \nabla \phi)$$

is element of the annihilator $j \in \mathcal{V}_0^\circ$ of $\mathcal{V}_0$ in $\mathcal{V}^* = H^{-1}(\Omega)$

$$\mathcal{V}_0^\circ := \{l \in H^{-1}(\Omega) \,\big|\, l(\phi) = 0 \; \forall \phi \in \mathcal{V}_0\}.$$

Proof. This is a direct consequence of (2.4). □

Now, for every $j \in \mathcal{V}_0^\circ$ we a looking for a pressure $p \in \mathcal{L}$. Surjectivity must therefore be shown with respect to the space $\mathcal{V}_0^\circ$. We can specify the adjoint operator of the weak gradient.

Lemma 2.13. Let $-\operatorname{grad} : \mathcal{L} \to H^{-1}(\Omega)$ be the weak gradient. Its adjoint is given by the divergence $\operatorname{div} = -\operatorname{grad}'$ as operator

$$\operatorname{div} : H_0^1(\Omega) \to \mathcal{L}.$$

Proof. This directly follows by

$$\langle -\operatorname{grad} p, \mathbf{v}\rangle = (p, \operatorname{div} \mathbf{v}) = (\operatorname{div} \mathbf{v}, p) = \langle \operatorname{div} \mathbf{v}, p\rangle \quad \forall \mathbf{v} \in H_0^1(\Omega)$$

and noting $[H^{-1}(\Omega)]' = H_0^1(\Omega)^d$ and $[\mathcal{L}^d]' = \mathcal{L}^d$. □

We now state the main result that will be proven in several steps:

Theorem 2.14 (Weak gradient operator)**.** Let $\Omega \subset \mathbb{R}^d$ be a domain with Lipschitz boundary. The weak gradient operator

$$-\operatorname{grad} : \mathcal{L} \to \mathcal{V}_0^\circ$$

is an isomorphism. This implies: it is bijective and it holds for all $p \in L^2(\Omega)$

$$\| -\operatorname{grad} p\|_{-1} \leqslant c_1 \|p\| \leqslant c_2 \| -\operatorname{grad} p\|_{-1} \tag{2.8}$$

with constants $c_1, c_2 > 0$ that depend on the domain Ω and the dimension d only.

Remark 2.15 (On the constants and boundness of the weak gradient operator)**.** One of the estimates in (2.8 is trivial and follows by using the definition of the H^{-1}-norm

$$\| - \operatorname{grad} p\|_{-1} := \sup_{\phi \in H_0^1(\Omega)} \frac{(p, \nabla \cdot \phi)}{\|\nabla \phi\|} \leqslant \|p\| \sup_{\phi \in H_0^1(\Omega)} \frac{\|\nabla \cdot \phi\|}{\|\nabla \phi\|} \leqslant \|p\|$$

and the estimate of Lemma 2.5 for functions with trace zero. Hence, $c_1 = 1$. This estimate shows that the weak gradient operator is a continuous operator.

The second constant c_2 is usually larger than one and its inverse $\gamma = 1/c_2$ will be an important measure in the analysis and numerical analysis of the Stokes and Navier-Stokes equations. We will later refer to γ as the *inf-sup constant*. △

Before proving the theorem we formulate the corallary that guarantees the existence of a unique solution to the Stokes equations.

Theorem 2.16 (Solution to the Stokes equations)**.** Let Ω be a domain with Lipschitz boundary, $\mathbf{f} \in H^{-1}(\Omega)$. There exists a unique solution to the variational Stokes problem

$$(\nabla \mathbf{v}, \nabla \phi) - (p, \nabla \cdot \phi) + (\nabla \cdot \mathbf{v}, \xi) = (\mathbf{f}, \phi) \quad \forall \{\phi, \xi\} \in \mathcal{V} \times \mathcal{L}$$

and it holds

$$\|\nabla \mathbf{v}\| + \gamma \|p\| \leqslant 3\|\mathbf{f}\|_{-1},$$

where $\gamma = 1/c_2$ with the constant $c_2 > 0$ from Theorem 2.8.

Proof. Lemma 2.7 shows the existence of a unique $\mathbf{v} \in \mathcal{V}_0 \subset \mathcal{V}$ solving the divergence free Stokes equations with the bound

$$\|\nabla \mathbf{v}\| \leqslant \|\mathbf{f}\|_{-1}.$$

For the functional

$$j(\phi) = (\nabla \mathbf{v}, \nabla \phi) - (\mathbf{f}, \phi)$$

it holds $j \in \mathcal{V}_0^\circ \subset H^{-1}(\Omega)$. Therefore, by Theorem 2.14 there exists a unique pressure $p \in \mathcal{L}$ which is solution of the equation $-\operatorname{grad} p = j$, which in turn is equivalent to

$$\langle -\operatorname{grad} p, \phi \rangle := (p, \nabla \cdot \phi) = (\nabla \mathbf{v}, \nabla \phi) - (\mathbf{f}, \phi) \quad \forall \phi \in \mathcal{V}.$$

Hence, the pair $\{\mathbf{v}, p\} \in \mathcal{V}_0 \times \mathcal{L} \subset \mathcal{V} \times \mathcal{L}$ is a solution to the Stokes equations. By (2.8) it holds with $\gamma := 1/c_2$

$$\begin{aligned}\gamma \|p\| \leqslant \| - \operatorname{grad} p\|_{-1} = \sup_{\phi \in H_0^1(\Omega)^d} \frac{(p, \nabla \cdot \phi)}{\|\nabla \phi\|} = \sup_{\phi \in H_0^1(\Omega)^d} \frac{(\nabla \mathbf{v}, \nabla \phi) - (\mathbf{f}, \phi)}{\|\nabla \phi\|} \\ \leqslant \|\nabla \mathbf{v}\| + \|\mathbf{f}\|_{-1} \leqslant 2\|\mathbf{f}\|_{-1},\end{aligned}$$

where we used the a priori estimate for the velocity, taken from Lemma 2.7. □

Next we develop the proof of Theorem 2.14. In Remark 2.15 we have shown a bound for the weak gradient. It remains to show the inverse of this estimate as well as injectivity and surjectivity of the weak gradient operator.

Remark 2.17 (Theorem 2.14 in the finite dimensional case)**.** A linear operator $T : V \to W$ in finite dimensional vector spaces can be considered as matrix $A : \mathbb{R}^m \to \mathbb{R}^n$ with adjoint $A^* = A^T$. The role of the annihilator is taken by the orthogonal complement of the kernel of the transposed matrix

$$\ker(A^*)^\circ = \ker(A^T)^\perp.$$

It hence holds with

$$\mathbb{R}^n = \operatorname{rg}(A) \oplus \ker(A^T)$$

that

$$\ker(A^T)^\perp = \operatorname{rg}(A).$$

This shows bijectivity of the map $A : \mathbb{R}^m \setminus \ker(A) \to \ker(A^T)^\perp$. △

Remark 2.18 (Injectivity of the gradient)**.** Given smooth functions $p \in C^1(\Omega) \setminus \mathbb{R}$, the injectivity of the gradient operator follows from the fact, that two functions with the same gradient differ by a constant only. △

Remark 2.19 (Pressure solution for high regularities)**.** We we assume, that the right hand side of $-\operatorname{grad} p = j$ has high regularity, i.e. $j \in H^1(\Omega)^d$. Then, we take the divergence of the pressure equation to find

$$-\operatorname{div}(\operatorname{grad} p) = \operatorname{div} j \quad \Rightarrow \quad -\Delta p = \operatorname{div} j.$$

If we prescribe boundary values it will be easy to solve this equation. For finding a pressure to the Stokes problem this argumentation is of little use: first, j will usually lack the desired regularity and second, the pressure does not carry natural boundary conditions. Any artificial condition will alter the result.

There is however a large class of numerical schemes for the approximation of the Stokes equations that is based on such a pressure-Poisson problem. One the challenges in the design of such schemes is the proper treatment of the boundary. △

Remark 2.20 (Bijectivity of the divergence)**.** Lemma 2.13 shows that the adjoint operator (to the weak gradient) is the divergence $\operatorname{div} : H_0^1(\Omega)^d \to \mathcal{L}$. If an operator is bijective so is its adjoint. It is therefore equivalent to show bijectivity of the divergence operator. For every $g \in \mathcal{L}$ we search

$$\mathbf{w} \in H_0^1(\Omega)^d : \quad \operatorname{div} \mathbf{w} = g. \tag{2.9}$$

For solving this equation we make the approach

$$-\mathbf{w} = \nabla q,$$

such that

$$-\operatorname{div}\nabla q = -\Delta q = g.$$

Considering the boundary condition $\mathbf{n}\cdot\nabla q = 0$ this problem has a unique solution $q \in H^1(\Omega)$. Then for $\mathbf{w}$ it holds

$$\mathbf{w} = -\nabla q \quad \Rightarrow \quad \mathbf{n}\cdot\mathbf{w} = \mathbf{n}\cdot\nabla q = 0 \text{ on } \partial\Omega.$$

Comparing to the original problem (2.9) reveals a slight difference: while we obtain $\mathbf{n}\cdot\nabla\mathbf{w} = 0$ we lack the corresponding condition in tangential direction $\mathbf{t}\cdot\nabla\mathbf{w} = 0$. To reach this condition we need to solve an *over-determined* Laplace problem

$$-\Delta q = g \text{ in } \Omega, \quad \mathbf{n}\cdot\nabla q = 0 \text{ and } \mathbf{t}\cdot\nabla q = 0 \text{ on } \partial\Omega,$$

which - in the general case - is not possible. △

The main tool to show surjectivity of the gradient is the *Closed Range Theorem*, see [33, 1] which we cite in the general case

Theorem 2.21 (Closed Range Theorem)**.** Let X and Y be Banach spaces, $T : X \to Y$ a linear operator with dual operator $T' : Y' \to X'$. The following conditions are equivalent

1. $\operatorname{rg}(T) \subset Y$ is closed,
2. $\operatorname{rg}(T') \subset X'$ is closed,
3. $\operatorname{rg}(T') = \ker(T)^\circ$.

In our setting, the operator T will be the divergence operator with adjoint T' the weak gradient. Then, Y is the Hilbert space $\mathcal{L}$ and X the Hilbert space $H_0^1(\Omega)^d$. We have already shown, that every possible right hand side j in $-\operatorname{grad} p = j$ will be member of the annihilator $j \in \mathcal{V}_0^\circ \subset H^{-1}(\Omega)$, see Lemma 2.12. Hence, if we are able to show either condition *1.* or condition *2.* of the Closed Range Theorem, condition *3.* will give surjectivity.

We will aim at showing closedness of $\operatorname{rg}(-\operatorname{grad}) = \operatorname{rg}(\operatorname{div}')$ in H^{-1}. Therefore we give a proof for the implication $2. \Rightarrow 3.$ of the Closed Range Theorem for a Hilbert space X. This proof is taken from Schweizer [34] and less complex than the complete proof of the Closed Range Theorem in the general case.

Proof. (*of the implication* $2. \Rightarrow 3.$ *of the Closed Range Theorem for Hilbert spaces* X).

First we note that $\operatorname{rg}(T) \subset \ker(T')^\circ$. Let $y' \in Y'$ and $x \in \ker(X)$

$$\langle T'y', x\rangle = \langle y', Tx\rangle = 0.$$

Hence, the range of T' is othorgonal on the kernel of T.

We show $\ker(T)^\circ \subset \mathrm{rg}(T')$ by contradiction. Assume, that $0 \neq x' \in \ker(T)^\circ \backslash \mathrm{rg}(T')$. As $\mathrm{rg}(T')$ is closed in X' the *Theorem of Hahn-Banach*[1] shows the existence of a functional $j : X' \to \mathbb{R}$ with representation $j \in X$ (as X is a Hilbert space) and

$$\langle j, x' \rangle = 1 \text{ and } j = 0 \text{ on } \mathrm{rg}(T').$$

For every such $Tj \in Y$ and arbitrary $y' \in Y'$ it holds

$$\langle y', Tj \rangle = \langle T'y', j \rangle = 0.$$

This however shows that $Tj = 0$ hence $j \in \ker(T)$ which is a contradiction to $\langle j, x' \rangle = 1$ as $x' \in \ker(T)^\circ$. □

To apply the Closed Range theorem to our situation we must show **closedness of -grad in $\mathbf{H^{-1}(\Omega)}$**. For this let $p_k \in \mathcal{L}$ such that $\mathbf{w}_k := -\operatorname{grad} p_k \in H^{-1}(\Omega)$ is a convergent sequence in H^{-1}, e.g.

$$\mathbf{w}_k = -\operatorname{grad} p_k \to \mathbf{w} \in H^{-1}(\Omega).$$

Of course $\mathbf{w}_k$ is a Cauchy sequence in $H^{-1}(\Omega)$

$$\|\mathbf{w}_k - \mathbf{w}_l\|_{H^{-1}(\Omega)} \to 0 \quad k, l \to \infty.$$

Remark 2.22 (Closedness at higher regularity)**.** To make things easy we shift this situation to higher regularity assuming that $p_k \in H^1(\Omega) \setminus \mathbb{R}$ and $\mathbf{w}_k \in L^2(\Omega)^d$ such that it would read

$$\|\mathbf{w}_k - \mathbf{w}_l\|_{L^2(\Omega)} = \|\nabla(p_k - p_l)\|_{L^2(\Omega)} \to 0 \quad k, l \to \infty.$$

In L^2 and $H^1_0(\Omega) \setminus \mathbb{R}$ the Poincaré inequality gives

$$\|p_k - p_l\| \leqslant c_p \|\nabla(p_k - p_l)\| = \|\mathbf{w}_k - \mathbf{w}_l\| \to 0,$$

such that p_k is Cauchy in $L^2(\Omega)^d$ which is a complete space such that a limit $p_k \to p \in L^2(\Omega)$ exists. △

To cast this remark into our setting we need an estimate similar to the Poincaré estimate. This is exactly the goal of the following discussion:

Lemma 2.23 (Lions-Poincaré Lemma)**.** Let Ω be a bounded domain with Lipschitz boundary. There exists a constant $c_{lp} > 0$ such that

$$\|q\|_{L^2(\Omega)} \leqslant c_{lp} \| - \operatorname{grad} q\|_{H^{-1}(\Omega)} \quad \forall q \in \mathcal{L} := L^2(\Omega) \setminus \mathbb{R}.$$

[1] The Theorem of Hahn-Banach says, that a linear functional $j_0 : X_0 \to \mathbb{R}$ on a subspace $X_0 \subset X$ can be continuously extended to a linear functional $j : X \to \mathbb{R}$ on the whole space with $j_{|X_0} = j_0$ and $\|j\|_{X \to \mathbb{R}} \leqslant \|j_0\|_{X_0 \to \mathbb{R}}$. A corollary (the Hahn-Banach separation Theorem) says, that for $x \in X \setminus X_0$ there exists a functional $j : X \to \mathbb{R}$ with $j(x) = 1$ and $j_{|X_0} = 0$.

The proof to this estimate will require some preparation. But having the Lions-Poincaré Lemma, remark 2.22 can be transferred to the weak gradient clearpage.

The Poincaré inequality in H^1 holds for functions with trace zero or functions with average zero. It's basis is the H^1-norm that is defined as

$$\|u\|^2_{H^1(\Omega)} := \|u\|^2_{L^2(\Omega)} + \|\nabla u\|^2_{L^2(\Omega)}.$$

Our first step will be to show a similar relation between the L^2 norm and the H^{-1} norm. This is less obvious and actually the main difficulty in all steps required for showing existence of the unique pressure in the Stokes equations.

Lemma 2.24 (Lions' Lemma)**.** Let $\Omega \subset \mathbb{R}^d$ be a bounded domain with Liptschitz boundary. There exists a constant $c_l > 0$ such that

$$\|q\|_{L^2(\Omega)} \leqslant c\big(\|q\|_{H^{-1}(\Omega)} + \|-\operatorname{grad} q\|_{H^{-1}(\Omega)}\big) \quad \forall q \in L^2(\Omega).$$

While the inverse of this estimate $\|q\|_{H^{-1}} + \|-\operatorname{grad} q\|_{H^{-1}} \leqslant c\|q\|_{L^2}$ follows by the definition of the H^{-1}-norm and the weak gradient, Lions' Lemma is difficult to proof [**?**, **?**]. We give a simplified version - again taken from Schweizer [34] - that is restricted to simple domains $\Omega = (0, 2\pi)^d$.

Proof. (*of Lions' Lemma for* $\Omega = (0, 2\pi)^d$.) (*i*) On the 2π-cube we can easily specify eigenvectors of the Laplace operator by

$$w_\alpha(x) = \exp(i\alpha \cdot x), \quad \alpha = (\alpha_1, \alpha_2) \in \mathbb{Z}^d,$$

as

$$\nabla w_\alpha(x) = i\alpha \exp(i\alpha \cdot x) = i\alpha w_\alpha(x), \quad -\Delta w_\alpha(x) = |\alpha|^2 w_\alpha(x),$$

such that $\lambda_\alpha = |\alpha|^2$ is the corresponding eigenvalue.[2] By w_α for $\alpha \in \mathbb{Z}^d$ an orthogonal basis of L^2 with $\|w_\alpha\| = \pi^{d/2}$ is given and every $q \in L^2(\Omega)$ can be written as

$$q(x) = \sum_{\alpha \in \mathbb{Z}^d} \underbrace{(q, w_\alpha)_{L^2(\Omega)}}_{=: q_\alpha \in \mathbb{C}} w_\alpha(x).$$

Given $q \in H^1(\Omega)$ it holds

$$\nabla q(x) = \sum_{\alpha \in \mathbb{Z}^d} i\alpha q_\alpha w_\alpha(x).$$

[2] We have shown, that the inverse of the Stokes operator is bounded with eigenvalues that have zero as only accumulation point. The same holds for the inverse of the Laplace operator. This finding fits to the currect configuration. If the inverse is bounded with zero as accumulation point, the Laplace operator itself is unbounded and its only accumulation point is infinity.

By this we can express the norms of $q \in L^2(\Omega)$ or $q \in H^1(\Omega)$ as

$$\begin{aligned}\|q\|^2_{L^2(\Omega)} &= \pi^{d/2} \sum_{\alpha\in\mathbb{Z}^d} |q_\alpha|^2,\\ \|\nabla q\|^2_{L^2(\Omega)} &= \pi^{d/2} \sum_{\alpha\in\mathbb{Z}^d} |\alpha|^2|q_\alpha|^2,\\ \|q\|^2_{H^1(\Omega)} &= \pi^{d/2} \sum_{\alpha\in\mathbb{Z}^d} (1+|\alpha|^2)|q_\alpha|^2.\end{aligned}$$

(*ii*) Next, let $q \in L^2(\Omega)$ and $\phi \in H^1(\Omega)$ both with representations in the basis w_k. Then, it holds by orthogonality (Parseval's identity)

$$\begin{aligned}(q,\phi)_{L^2(\Omega)} &= \sum_{\alpha\in\mathbb{Z}^d} q_\alpha\phi_\alpha(w_\alpha,w_\alpha) = \pi^{d/2}\sum_{\alpha\in\mathbb{Z}^d} q_\alpha\phi_\alpha\\ &= \pi^{d/2}\sum_{\alpha\in\mathbb{Z}^d} (1+|\alpha|^2)^{-1/2} q_\alpha (1+|\alpha|^2)^{1/2}\phi_\alpha\\ &\leqslant \left(\pi^d \sum_{\alpha\in\mathbb{Z}^d} (1+|\alpha|^2)^{-1}|q_\alpha|^2\right)^{\frac12} \underbrace{\left(\pi^d \sum_{\alpha\in\mathbb{Z}^d} (1+|\alpha|^2)|\phi_\alpha|^2\right)^{\frac12}}_{=:\|\phi\|_{H^1(\Omega)}}.\end{aligned}$$

We conclude

$$|(q,\phi)_{L^2(\Omega)} \leqslant \left(\pi^d \sum_{\alpha\in\mathbb{Z}^d} \frac{|q_\alpha|^2}{1+|\alpha|^2}\right)^{\frac12} \|\phi\|_{H^1(\Omega)}$$

For $\phi = q$ this relation holds as an equality. Hence we get an expression for the H^{-1}-norm:

$$\|q\|_{H^{-1}(\Omega)} := \sup_{\phi\in H^1(\Omega)} \frac{|(q,\phi)|}{\|\phi\|_{H^1(\Omega)}} = \left(\pi^d \sum_{\alpha\in\mathbb{Z}^d} \frac{|q_\alpha|^2}{1+|\alpha|^2}\right)^{\frac12}.$$

(*iii*) Now, let $u \in H^1(\Omega)$. It holds

$$\begin{aligned}\|u\|^2_{L^2(\Omega)} &= \pi^{d/2}\sum_{\alpha\in\mathbb{Z}^d} |u_\alpha|^2 = \pi \sum_{\alpha\in\mathbb{Z}^d} \frac{|u_\alpha|^2}{1+|\alpha|^2} + \pi \sum_{\alpha\in\mathbb{Z}^d} \frac{1}{1+|\alpha|^2}|\alpha|^2\cdot|u_\alpha|^2\\ &= \|u\|^2_{H^{-1}(\Omega)} + \|\nabla u\|_{H^{-1}(\Omega)}\end{aligned}$$

To extend this result to functions $q \in L^2(\Omega)$ let $u_k \in H^1(\Omega)$ be a sequence with $u_k \to q \in L^2(\Omega)$. Then it holds $\nabla u_k \to \nabla q$ in $H^{-1}(\Omega)$ (as the gradient is a continuous operator). Then it holds

$$\begin{aligned}\|q\| &\leqslant \|q-u_k\| + \|u_k\| = \|q-u_k\| + \|u_k\|_{H^{-1}(\Omega)} + \|\nabla u_k\|_{H^{-1}(\Omega)}\\ &= \|q-u_k\| + \|q\|_{H^{-1}(\Omega)} + \|\nabla q\|_{H^{-1}(\Omega)} + \|q-u_k\|_{H^{-1}(\Omega)} + \|\nabla(q-u_k)\|_{H^{-1}(\Omega)}\\ &\to \|q\|_{H^{-1}(\Omega)} + \|\nabla q\|_{H^{-1}(\Omega)} \quad (k\to\infty).\end{aligned}$$

□

Lions' lemma will give us an embedding relation similar to $H^1 \hookrightarrow L^2$ in spaces of lower regularity.

Lemma 2.25 (Compact embedding in Sobolev spaces with negative exponent)**.** Let $\Omega \subset \mathbb{R}^d$ be a bounded domain with Lipschitz boundary. The embedding $L^2(\Omega) \hookrightarrow H^{-1}(\Omega)$ is compact.

Proof. (*This proof is taken from* [34]) We assume that this lemma does not hold. Then there exists a sequence $u_k \in L^2(\Omega)$ that is bounded in L^2. Then there exists a weakly convergent subsequence, denoted again by $u_k \rightharpoonup u \in L^2(\Omega)$. By considering $u_k - u$ we can assume, that $u_k \rightharpoonup 0$.

Now we assume that u_k does *not converge strongly* in $H^{-1}(\Omega)$ to zero. Then, there exists a sequence $\phi_k \in H^1(\Omega)$ of bounded test functions with

$$\langle u_k, \phi_k \rangle_{H^{-1}(\Omega)\times H^1(\Omega)} = 1.$$

As the embedding $H^1(\Omega) \hookrightarrow L^2(\Omega)$ is compact there exists a converging (in L^2) subsequence with $\phi_{k'} \to \phi$ in $L^2(\Omega)$ for this subsequence it holds

$$1 = \langle u_{k'}, \phi_{k'} \rangle_{H^{-1}(\Omega)\times H^1(\Omega)} = (u_{k'}, \phi_{k'}) \to (u, \phi) = 0,$$

which is a contradiction. □

With this embedding result we can give a proof to Lions-Poincaré Lemma, lemma 2.23.

Proof. to Lemma 2.23. We aim at prooving the estimate

$$\|q\|_{L^2(\Omega)} \leqslant c_{pl} \| - \operatorname{grad} q\|_{H^{-1}(\Omega)} \quad \forall q \in \mathcal{L} := L^2(\Omega) \setminus \mathbb{R}. \tag{2.10}$$

We give the proof by contradiction and assume that a sequence $q_k \in \mathcal{L} \subset L^2(\Omega)$ exists with $\|q_k\|_{L^2(\Omega)} = 1$ but $-\operatorname{grad} q_k \to 0$ in $H^{-1}(\Omega)$. As q_k is bounded in L^2 it has a weakly convergent subsequence $q_k \rightharpoonup \tilde{q} \in L^2(\Omega)$. We want to show that $\tilde{q} = 0$. We know that $\nabla q_k \to \nabla \tilde{q} = 0$ in $H^{-1}(\Omega)$. Hence $\tilde{q} \in \mathcal{L}$ has average zero and it's weak gradient is zero. Therefore $\tilde{q} = 0$.

Lemma 2.25 says that there exists H^{-1}-converging subsequence (also denoted by q_k) with $\|q_k\|_{H^{-1}(\Omega)} \to 0$. Then, Lemma 2.24 gives

$$\|q_k\| \leqslant c\big(\|q_k\|_{H^{-1}} + \| - \operatorname{grad} q_k\|_{H^{-1}}\big) \to 0$$

in contradiction to $\|q_k\| = 1$ such that estimate (2.10) holds. □

This leads us to the range of the weak gradient operator.

Lemma 2.26 (Closed range of the weak gradient)**.** Let $\Omega \subset \mathbb{R}^d$ be a bounded domain with Lipschitz boundary. The weak gradient operator

$$-\operatorname{grad} : \mathcal{L} \to H^{-1}(\Omega)$$

has a closed range

$$\operatorname{rg}(-\operatorname{grad}) \subset H^{-1}(\Omega).$$

Proof. Let $q_k \in L^2(\Omega)$ be a sequence that is convergent in $H^{-1}(\Omega)$. Hence Lions-Poincaré lemma says

$$\|q_k - q_l\|_{L^2(\Omega)} \leqslant c_{lp}\| - \operatorname{grad}(q_k - q_l)\|_{H^{-1}(\Omega)} \to 0 \quad (k, l \to \infty),$$

such that q_k is a Cauchy sequence in $L^2(\Omega)$ with limit $q_k \to q \in L^2(\Omega)$. Since the gradient operator is continuous it holds

$$-\operatorname{grad}\ q_k \to -\operatorname{grad}\ q \in H^{-1}(\Omega)$$

with $-\operatorname{grad}\ q \in \operatorname{rg}(-\operatorname{grad})$. □

Finally, the closedness of the divergence operator allows us to apply the Closed Range theorem (Theorem 2.21). We find that

$$\operatorname{rg}(-\operatorname{grad}) = \ker(\operatorname{div})^\circ = \mathcal{V}_0^\circ.$$

The weak gradient is a surjection on the annihilator of $\mathcal{V}_0$ in $H^{-1}(\Omega)$. Therefore for every possible right hand side $l \in \mathcal{V}_0^\circ$ we find a pressure $p \in \mathcal{L}$ solving the Stokes equations.

It remains to formally injectivity of the weak gradient and an estimate for $\|p\|$.

Lemma 2.27 (Injectivity of the weak gradient)**.** The weak gradient operator $-\operatorname{grad} : \mathcal{L} \to \mathcal{V}_0^\circ$ is injective and it holds

$$\|p\|_{L^2(\Omega)} \leqslant c_{lp}\| - \operatorname{grad} p\|_{H^{-1}(\Omega)} \leqslant c_{lp}\|p\|_{L^2(\Omega)} \quad \forall p \in \mathcal{L}.$$

Proof. The first estimate is exactly Lions-Poincaré lemma. The second estimate follows by the definition of the H^{-1}-norm in $H_0^1(\Omega)^d$ as

$$\| - \operatorname{grad}\ p\|_{H^{-1}(\Omega)} := \sup_{\phi \in H_0^1(\Omega)} \frac{\langle -\operatorname{grad} p, \phi\rangle}{\|\nabla\phi\|} = \sup_{\phi \in H_0^1(\Omega)} \frac{(p, \operatorname{div}\phi)}{\|\nabla\phi\|} \leqslant \sup_{\phi \in H_0^1(\Omega)} \frac{\|p\|\,\|\operatorname{div}\phi\|}{\|\nabla\phi\|}.$$

As we consider homogenous Dirichlet values, Lemma 2.6 gives the estimate

$$\| - \operatorname{grad}\ p\|_{H^{-1}(\Omega)} \leqslant \|p\|_{L^2(\Omega)}.$$

Now assume, that $p_1, p_2 \in \mathcal{L}$ are two pressure solutions to the right hand side $l \in \mathcal{V}_0^\circ$. For $q := p_1 - p_2 \in \mathcal{L}$ it holds with the Lions-Poincaré lemma

$$\|q\| \leqslant \| - \operatorname{grad}\ q\|_{H^{-1}(\Omega)} = \| - \operatorname{grad}(p_1 - p_2)\|_{H^{-1}(\Omega)} = \|l - l\|_{H^{-1}(\Omega)} = 0.$$

Hence the pressure is unique. □

2.1.4 The inf-sup condition

We end with a summary of this lengthy argumentation. The proof to a unique solution to the Navier-Stokes equation is split into two parts. First we show the existence of a velocity, second the existence of a corresponding pressure.

1. By restricting the variational formulation to the space of divergence free functions

$$\mathcal{V}_0 := \{\phi \in H_0^1(\Omega)^d \mid (\operatorname{div}\phi, \xi) = 0 \quad \forall \xi \in \mathcal{L}\}$$

   it corresponds to the vector-Laplace. By showing that $\mathcal{V}_0$ is closed subspace of $\mathcal{V}$ we are in the setting of Hilbert spaces and a unique solution satisfying the bound

$$\|\nabla \mathbf{v}\| \leqslant \|\mathbf{f}\|_{H^{-1}}$$

   is obtained with Riesz respresentation theorem.

2. The corresponding pressure problem is formulated in a weak setting

$$-\operatorname{grad} p = l, \quad l(\phi) := (\mathbf{f}, \phi) - (\nabla \mathbf{v}, \nabla \phi) \quad \forall \phi \in \mathcal{V},$$

   where $\mathbf{v}$ is the velocity solution. The right hand side is element of the annihilitor of $\mathcal{V}_0$ in $H^{-1}(\Omega)$, i.e. $l \in \mathcal{V}_0^\circ$. We show that the weak gradient is an isomorphism

$$-\operatorname{grad} : \mathcal{L} \to \mathcal{V}_0^\circ.$$

   To show surjectivity we use the Closed Range theorem. To proof that the range of the divergence (which is the dual to the weak gradient) is closed we need the Lions-Poincaré estimate

$$\|p\|_{L^2} \leqslant c_{lp} \| -\operatorname{grad} p\|_{H^{-1}(\Omega)} \quad \forall p \in \mathcal{L},$$

   which corresponds to the classical Poincaré estimate - but which is stated in function spaces of lower regularity. The proof of this estimate is nontrivial. The Lions-Poincaré estimate also gives uniqueness and the bound for the pressure.

Although the existence theory for the Stokes problem is finished at this point we formulate another lemma which gives equivalent formulations to Theorem 2.14 which says that the weak gradient is an isomorphism.

Theorem 2.28 (inf-sup condition)**.** The following three properties are equivalent

(i) The weak gradient operator $-\operatorname{grad} : \mathcal{L} \to \mathcal{V}_0^\circ$ is an isomorphism.

(ii) For every $p \in L^2(\Omega)$ it holds

$$\gamma \|p\| \leqslant \|\operatorname{grad} p\|_{-1} \quad \forall p \in \mathcal{L}, \tag{2.11}$$

   where $\gamma > 0$ is a constant.

(iii) The inf-sup condition holds

$$\inf_{\xi\in\mathcal{L}} \sup_{\phi\in\mathcal{V}} \frac{(\xi, \nabla\cdot\phi)}{\|\xi\|\,\|\nabla\phi\|} = \gamma > 0, \tag{2.12}$$

with $\gamma > 0$ from (2.11).

Proof. This theorem goes back to Nečas.

a) (i)⇒(ii) Condition (i) says that for every $p \in \mathcal{L}$ there exists a unique $l \in \mathcal{V}_0^\circ$ with $-\operatorname{grad} p = l$ that satisfies the bound (2.11).

b) (ii)⇔(iii) By the definition of the H^{-1}-norm condition (ii) reads

$$\gamma\|p\| \leqslant \sup_{\phi\in H_0^1(\Omega)} \frac{\langle -\operatorname{grad} p, \phi\rangle}{\|\nabla\phi\|} = \sup_{\phi\in H_0^1(\Omega)} \frac{(p, \nabla\phi)}{\|\nabla\phi\|}$$

As this condition holds for all $p \in \mathcal{L}$ it is equivalent to

$$\gamma = \sup_{\phi\in H_0^1(\Omega)} \frac{(p, \nabla\phi)}{\|\nabla\phi\|\,\|p\|} \quad \forall p \in \mathcal{L} \quad \Leftrightarrow \quad \gamma \leqslant \inf_{p\in\mathcal{L}} \sup_{\phi\in H_0^1(\Omega)} \frac{(p, \operatorname{div}\phi)}{\|\nabla\phi\|\,\|p\|}.$$

We take $\gamma > 0$ as the supremum.

c) (iii)⇒(i) We know that (iii) and (ii) are equivalent. Injectivity follows by (2.11). Further the bounds in both directions follow from (2.11) and the definition of the H^{-1} norm. It remains to show the surjectivity of the gradient. This is again accomplished by using the Closed Range theorem. Closedness of the gradient is shown with help of estimate (2.11) as we have argued before. □

This theorem gives various version of a condition that is equivalent to Lions-Poincaré estimate. The inf-sup condition will be of great use in the numerical analysis of the Stokes equations. When discussing finite element discretizations of the equation we will come across a discrete version of this estimate. The *inf-sup condition* allows for several similar reformulation that are useful in various contexts, such as:

Corollary 2.29. For every $p \in \mathcal{L}$ there exists a $\mathbf{v} \in \mathcal{V}$ with $\|\nabla\mathbf{v}\| = 1$ such that

$$\frac{\gamma}{2}\|p\| \leqslant (p, \operatorname{div}\mathbf{v}).$$

Proof. The *inf-sup condition* gives

$$\gamma\|p\| \leqslant \sup_{\substack{\phi\in\mathcal{V}_0\\ \|\nabla\phi\|=1}} (p, \operatorname{div}\phi).$$

We consider a sequence $\mathbf{v}_k \in \mathcal{V}_0$ with

$$(p, \operatorname{div} \mathbf{v}_k) \to \sup_{\substack{\phi \in \mathcal{V}_0 \\ \|\nabla \phi\|=1}} (p, \operatorname{div} \phi) \geqslant \gamma \|p\|$$

For some $k_0 \in \mathbb{N}$ it will hold

$$(p, \operatorname{div} \mathbf{v}_k) \geqslant \frac{\gamma}{2}\|p\| \quad \forall k > k_0.$$

We take such a $\mathbf{v}_k$. □

2.1.5 Stokes as a coercive system

The inf-sup condition can also be used to treat the full Stokes system in saddle-point form at once and to show existence of a solution.

Lemma 2.30 (Stokes as a coercive system)**.** Let

$$U := (\mathbf{v}, p) \in \mathcal{X} := \mathcal{V} \times \mathcal{L}$$

and $A : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ be defined as

$$A(U, \Phi) := (\nabla \mathbf{v}, \nabla \phi) - (p, \operatorname{div} \phi) + (\operatorname{div} \mathbf{v}, \xi) \quad \forall \Phi := (\phi, \xi) \in \mathcal{X}.$$

The bilinear form $A(\cdot, \cdot)$ is continues and coercive

$$|A(U, \Phi)| \leqslant |\!|\!| U |\!|\!| \, |\!|\!| \Phi |\!|\!|, \qquad \sup_{\substack{\Phi \in \mathcal{X} \\ |\!|\!| \Phi |\!|\!|=1}} A(U, \Phi) \geqslant c |\!|\!| U |\!|\!|$$

with a constant $c > 0$ and where

$$|\!|\!| U |\!|\!| := \left(\|\nabla \mathbf{v}\|^2 + \gamma^2 \|p\|^2 \right)^{\frac{1}{2}}$$

defines a norm on $\mathcal{X}$.

Proof. (i) To show coercivity we use a common technique in analysis of variational formulations and also in the numerical analysis of finite element methods. We construct the estimate by testing with apropriate test-functions. First, for given $U = (\mathbf{v}, p)$ we choose

$$\Phi_1 = U$$

to get

$$A(U, \Phi_1) = (\nabla \mathbf{v}, \nabla \mathbf{v}) - (p, \operatorname{div} \mathbf{v}) + (\operatorname{div} \mathbf{v}, p) = \|\nabla \mathbf{v}\|^2.$$

Next, corollary (2.29) shows us the existence of a $\tilde{\mathbf{v}} \in \mathcal{V}_0$ with $\|\nabla \mathbf{v}\| = 1$ and

$$\frac{\gamma}{2}\|p\| = (\operatorname{div} \tilde{\mathbf{v}}, p).$$

We test with

$$\Phi_2 = (-\tilde{\mathbf{v}}\|p\|, 0)$$

to get

$$\begin{aligned} A(U, \Phi_2) &= -(\nabla \mathbf{v}, \nabla \tilde{\mathbf{v}})\|p\| + (p, \operatorname{div} \tilde{\mathbf{v}})\|p\| \\ &= -(\nabla \mathbf{v}, \nabla \tilde{\mathbf{v}})\|p\| + \frac{\gamma}{2}\|p\|^2. \qquad \geqslant -\|\nabla \mathbf{v}\| \underbrace{\|\nabla \tilde{\mathbf{v}}\|}_{=1} \|p\| + \frac{\gamma}{2}\|p\|^2. \end{aligned}$$

We apply Young's inequality

$$A(U, \Phi_2) \geqslant -\frac{1}{\gamma}\|\nabla \mathbf{v}\|^2 + \frac{\gamma}{4}\|p\|^2 + \frac{\gamma}{2}\|p\|^2 = \frac{\gamma}{4}\|p\|^2 - \frac{1}{\gamma}\|\nabla \mathbf{v}\|^2.$$

We combine both test-functions to $\Phi := \Phi_1 + \frac{\gamma}{2}\Phi_2$ to get

$$A(U, \Phi) \geqslant \frac{1}{2}\|\nabla \mathbf{v}\|^2 + \frac{\gamma^2}{8}\|p\|^2 \geqslant \frac{1}{8}|||U|||^2.$$

Finally, we show

$$|||\Phi||| \leqslant |||\Phi_1||| + \frac{\gamma}{2}|||\Phi_2||| \leqslant |||U||| + \frac{\gamma}{2} \underbrace{\|\nabla \tilde{\mathbf{v}}\|}_{=1} \|p\| \leqslant \frac{3}{2}|||U|||$$

by which we show the coercivity estimate.

(*ii*) The continuity follows by Cauchy-Schwarz estimate

$$A(U, \Phi) \leqslant \|\nabla \mathbf{v}\| \, \|\nabla \phi\| + \|p\| \, \|\nabla \cdot \phi\| + \|\operatorname{div} \mathbf{v}\| \, \|\xi\|$$

and the estimate for the divergence $\|\operatorname{div} \mathbf{v}\| \leqslant \|\nabla \mathbf{v}\|$ in $H_0^1(\Omega)^d$

$$A(U, \Phi) \leqslant \big(\|\nabla \mathbf{v}\| + \|p\|\big)\big(\|\nabla \phi\| + \|\xi\|\big) \leqslant c(\gamma)|||U||| \, |||\Phi|||.$$

□

Coercivity and continuity in $\mathcal{X}$ can be used to show existence of a unique solution by Lax-Milgram.

Despite the special saddle-point character of the Stokes equations it shows that we still get a unique solution that continuously depends on the right hand side $\mathbf{f}$. We only get L^2-regularity for the pressure. The most important tool in the analysis of incompressible flows is the inf-sup condition. If the right hand side $\mathbf{f}$ and the domain is sufficiently regular, we will get higher regularity of the solution. Here, the same rule of thump holds as for the Laplace equation:

Lemma 2.31 (Regularity of the Stokes solution)**.** Let Ω be a convex polygonal domain and $\mathbf{f} \in L^2(\Omega)^d$. Then the solution of the Stokes equations is bounded

$$\|\nabla^2 \mathbf{v}\| + \|\nabla p\| \leqslant c_s \|\mathbf{f}\|,$$

with a stability constant $c_s > 0$.

If $\Omega \subset \mathbb{R}^d$ is a domain with smooth C^{k+2}-boundary for $k \geqslant 0$ and $\mathbf{f} \in H^k(\Omega)^d$ it holds

$$\|\mathbf{v}\|_{H^{k+2}(\Omega)} + \|p\|_{H^{k+1}(\Omega)} \leqslant c \|\mathbf{f}\|_{H^k(\Omega)}.$$

Proof. For a proof to these results, we refer to the literature [38, 13]. □

2.2 Existence and uniqueness for the Navier-Stokes Equations

2.2.1 The stationary Navier-Stokes equations

Next, we discuss the stationary Navier-Stokes equations including the nonlinearity

$$\{\mathbf{v}, p\} \in \mathcal{V} \times L, \quad \mathcal{V} := H_0^1(\Omega; \partial\Omega)^d, \quad \mathcal{L} := L^2(\Omega) \setminus \mathbb{R} :$$
$$\frac{1}{Re}(\nabla \mathbf{v}, \nabla \phi) + (\mathbf{v} \cdot \nabla \mathbf{v}, \phi) - (p, \nabla \cdot \phi) + (\nabla \cdot \mathbf{v}, \xi) = (\mathbf{f}, \phi)$$
$$\forall \{\phi, \xi\} \in \mathcal{V} \times \mathcal{L}, \quad (2.13)$$

again considering homogenous Dirichlet conditions $\mathbf{v} = 0$ only. Here, this restriction is essential not merely given for technical reasons, as the following Lemma shows:

Lemma 2.32 (Nonlinearity of the Navier-Stokes equations)**.** For $\mathbf{v}, \mathbf{w} \in H_0^1(\Omega)^d$ with $\operatorname{div} \mathbf{v} = 0$ it holds:

$$(\mathbf{v} \cdot \nabla \mathbf{w}, \mathbf{w}) = 0. \quad (2.14)$$

In the case of an outflow boundary $\Gamma^{out} \subset \partial\Omega$ it holds for all $\mathbf{v}, \mathbf{w} \in H_0^1(\Omega; \Gamma^D)^d$ with $\operatorname{div} \mathbf{v} = 0$

$$\Big((\mathbf{v} \cdot \nabla)\mathbf{w}, \mathbf{w}\Big) = \frac{1}{2} \int_{\Gamma^{out}} \mathbf{n} \cdot \mathbf{v} |\mathbf{w}|^2 \, ds. \quad (2.15)$$

Proof. In the case of general boundary conditions it holds

$$\begin{aligned}
\Big((\mathbf{v} \cdot \nabla)\mathbf{w}, \mathbf{w}\Big)_\Omega &= \sum_{i,j} (v_j \partial_j w_i, w_i)_\Omega \\
&= \sum_{i,j} \Big\{ \int_{\partial\Omega} n_j w_i v_j w_i \, ds - (w_i, \partial_j v_j w_i)_\Omega - (w_i, v_j \partial_j w_i)_\Omega \Big\} \\
&= -\underbrace{(\mathbf{w}, (\operatorname{div} \mathbf{v})\mathbf{w})_\Omega}_{=0} - ((\mathbf{v} \cdot \nabla)\mathbf{w}, \mathbf{w})_\Omega + \int_{\partial\Omega} (\mathbf{n} \cdot \mathbf{v}) |\mathbf{w}|^2 \, ds.
\end{aligned}$$

This shows the two assertions. □

This special structure of the nonlinearity will be the key to theoretical analysis of the incompressible Navier-Stokes equations.

Lemma 2.33 (Stability estimate for the velocity)**.** Let $\mathbf{v} \in \mathcal{V}_0 \subset H_0^1(\Omega)^d$ be a velocity field solving the Navier-Stokes equations. It holds for $\mathbf{f} \in L^2(\Omega)^d$

$$\|\nabla \mathbf{v}\| \leqslant \nu^{-1} \|\mathbf{f}\|_{-1}.$$

Proof. This results immediately follows with Lemma 2.32. □

Remark 2.34 (Outflow conditions and stability estimates)**.** Lemma 2.32 shows that the nonlinearity of the Navier-Stokes equations is only controllable, if Dirichlet or at least no-penetration conditions

$$\mathbf{v} \cdot \mathbf{n} = 0,$$

are given on all boundaries. For the *do-nothing* conditions but also for the *no-stress* condition introduced in Section 1.4.2 a boundary term remains. The problem of this remaining boundary term

$$\frac{1}{2} \int_{\Gamma^{out}} \mathbf{n} \cdot \mathbf{n} |\mathbf{w}|^2 \, do,$$

is the unknown sign. If there would be only outflow, i.e. $\mathbf{n} \cdot \mathbf{v} \geqslant 0$, we still get stability in the sense of Lemma 2.33. In the general setting, the boundary term however can be negative or positive. Braack and Mucha [4] introduced a modification of the do-nothing condition, denoted the *directional do-nothing condition* that cancels the negative part of the boundary term and results in

$$-p\mathbf{n} + \rho\nu \mathbf{n} \cdot \nabla \mathbf{v} - \frac{1}{2} (\mathbf{v} \cdot \mathbf{n})_- \mathbf{v} = 0 \text{ on } \Gamma^{out},$$

where by $(\mathbf{v} \cdot \mathbf{n})_-$ we denote

$$(\mathbf{v} \cdot \mathbf{n})_- = \begin{cases} 0 & \mathbf{v} \cdot \mathbf{n} \geqslant 0, \\ \mathbf{v} \cdot \mathbf{n} & \mathbf{v} \cdot \mathbf{n} < 0. \end{cases}$$

This condition is easily realized by a modification of the variational formulation

$$(\mathbf{v} \cdot \nabla \mathbf{v}, \phi) + (\rho\nu \nabla \mathbf{v}, \nabla \phi) - (p, \nabla \cdot \phi) - \frac{1}{2} \int_{\Gamma^{out}} (\mathbf{v} \cdot \mathbf{n})_- \mathbf{v} \cdot \phi \, do = (\mathbf{f}, \phi).$$

Braack and Mucha can show existence and uniqueness of solutions (for small data). Furthermore, they report better numerical stability when using this directional do-nothing condition. Finally, this modified condition still allows for Poiseuille and Couette flow as well as vortices to leave the domain with little impact. See [4] for details. △

Like for the Stokes equations, proofs for existence and uniqueness are split into first finding the velocity (this is a nonlinear problem now) and second, finding an appropriate pressure. While this second part is exactly as for the linear Stokes problem, showing existence and uniqueness of a velocity requires careful treatment of the nonlinearity.

$$\nu(\nabla \mathbf{v}, \nabla \phi) + ((\mathbf{v} \cdot \nabla)\mathbf{v}, \phi) = (\mathbf{f}, \phi) \quad \forall \phi \in \mathcal{V}_0. \tag{2.16}$$

Theorem 2.35 (Solutions to the Navier-Stokes equations)**.** Let $\Omega \subset \mathbb{R}^d$ be a domain with Lipschitz boundary. Further, let $\mathbf{f} \in H^{-1}(\Omega)$. There exists a solution $\{\mathbf{v}, p\} \in \mathcal{V} \times \mathcal{L}$ to the Navier-Stokes equations (2.13) for every Reynolds number. It holds

$$\|\nabla \mathbf{v}\| + \|p\| \leqslant c\|\mathbf{f}\|_{-1}.$$

This solution is unique, if

$$c^2 \nu^{-2} \|\mathbf{f}\|_{-1} \leqslant 1,$$

where $c > 0$ is a constant depending on the domain Ω.

Proof. The proof will be split into several parts. Again, velocity and pressure can be handled separately. First, by a restriction of the equation to the space of divergence-free functions, we are able to solve the velocity problem. Next, a corresponding pressure is found with help of the inf-sup condition.

Due to the nonlinearity we cannot directly use the common theorems from linear functional analysis like Riesz or Lax-Milgram. Instead we will introduce a Galerkin approach in step (*v.1*) using finite dimensional (but nonlinear) problems. In step (*v.2*) we will show convergence of a finite dimensional fixed-point iteration. In step (*v.3*) we show that the sequence of finite dimensional solutions convergens to a solution $\mathbf{v} \in \mathcal{V}_0$ of the Navier-Stokes equations. Step (*v.4*) will show uniqueness under the stated stronger conditions. Finally, in Step (*p*) the corresponding pressure solution will be constructed. Given the previous discussion of the Stokes equations, this step will be easy.

(*v.1*) *Galerkin-approach:* We construct finite dimensional subspaces of $\mathcal{V}_0$:

$$\mathcal{V}_m := \operatorname{span}\{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_m\},$$

where the $\mathbf{w}_k$ are the orthonormal Eigenfunctions of the Stokes operator, see Section 2.1. Orthonormality or even the property of being Eigenfunctions is not required for this proof. It is however essential that the union $\bigcup_{m \geqslant 1} \mathcal{V}_m$ is dense in $\mathcal{V}_0$, such that every $\mathbf{v} \in \mathcal{V}_0$ can be approximated by a sequence $\mathbf{v}_m \in \mathcal{V}_m$ for $m \to \infty$.

For $m \geqslant 1$ we define the finite dimensional problem

$$\mathbf{v}_m \in \mathcal{V}_m : \quad \nu(\nabla \mathbf{v}_m, \nabla \phi_m) + ((\mathbf{v}_m \cdot \nabla)\mathbf{v}_m, \phi_m) = (\mathbf{f}, \phi_m) \quad \forall \phi_m \in \mathcal{V}_m. \tag{2.17}$$

To cope with the nonlinearity, we introduce a fixed-point map $Q_m : \mathcal{V}_m \to \mathcal{V}_m$ by:

$$\nu(\nabla Q_m(\mathbf{v}_m), \nabla \phi_m) + ((\mathbf{v}_m \cdot \nabla) Q_m(\mathbf{v}_m), \phi_m) = (\mathbf{f}, \phi_m) \quad \forall \phi_m \in \mathcal{V}_m. \tag{2.18}$$

Injectivity of this linear and finite dimensional problem induces bijectivity. For the homogenous equation it follows by diagonal testing with help of (2.14) that

$$\|\nabla Q_m(\mathbf{v}_m)\|^2 + \underbrace{(\mathbf{v}_m \cdot \nabla Q_m(\mathbf{v}_m), Q_m(\mathbf{v}_m))}_{=0} = 0 \quad \Rightarrow \quad Q_m(\mathbf{v}_m) = 0.$$

Problem (2.18) has a unique solution $Q_m(\mathbf{v}_m) \in V_m$ for every $\mathbf{f} \in H^{-1}$ and $\mathbf{v}_m \in \mathcal{V}_m$. For this solution we get the following estimate

$$\nu\|\nabla Q_m(\mathbf{v}_m)\|^2 \leqslant \|\mathbf{f}\|_{-1}\|\nabla Q_m(\mathbf{v}_m)\| \quad \Rightarrow \quad \|\nabla Q_m(\mathbf{v})\| \leqslant \nu^{-1}\|\mathbf{f}\|_{-1}. \tag{2.19}$$

(v.2) Fixed-point argument: We will show that the mapping $Q_m : \mathcal{V}_m \to \mathcal{V}_m$ is fixed-point mapping $Q_m(\mathbf{v}) = \mathbf{v}$ by using Brouwer's fixed-point theorem: every continuous mapping of a compact and convex subset of a Banachspace into itself has at least one fixpoint. We define

$$B_R := \{\phi \in \mathcal{V}_m,\ \|\nabla\phi\| \leqslant R := \nu^{-1}\|f\|_{-1}\},$$

and show that $Q_m : B_R \to B_R$ is a continuous mapping into itself. For the solution of (2.18) if follows by (2.19) for $\mathbf{v} \in B_R$ that

$$\|\nabla Q_m(\mathbf{v})\| \leqslant \nu^{-1}\|\mathbf{f}\|_{-1} =: R \quad \Rightarrow \quad Q_m(\mathbf{v}) \in B_R,$$

such that $Q_m : B_R \to B_R$ is a mapping into itself.

Next, we show that Q_m is Lipschitz and hence also continuous. Let $\mathbf{v}, \mathbf{w} \in \mathcal{V}_m$:

$$\begin{aligned} 0 &= \nu(\nabla(Q_m(\mathbf{v}) - Q_m(\mathbf{w})), \nabla\phi) + ((\mathbf{v}\cdot\nabla)Q_m(\mathbf{v}) - (\mathbf{w}\cdot\nabla)Q_m(\mathbf{w}), \phi) \\ &= \frac{1}{Re}(\nabla(Q_m(\mathbf{v}) - Q_m(\mathbf{w})), \nabla\phi) + (((\mathbf{v}-\mathbf{w})\cdot\nabla)Q_m(\mathbf{v}), \phi) \\ &\qquad + ((\mathbf{w}\cdot\nabla)(Q_m(\mathbf{v}) - Q_m(\mathbf{w})), \phi) \end{aligned}$$

For $\phi := Q_m(\mathbf{v}) - Q_m(\mathbf{w})$ and using (2.14) it holds:

$$\nu\|\nabla(Q_m(\mathbf{v}) - Q_m(\mathbf{w}))\|^2 = -((\mathbf{v}-\mathbf{w})\cdot\nabla Q_m(\mathbf{v}), Q_m(\mathbf{v}) - Q_m(\mathbf{w})).$$

The product on the right hand side can be estimated using the generalized Hölder's inequality

$$\|fgh\|_{L^1} \leqslant \|f\|_{L^{p_1}}\|g\|_{L^{p_2}}\|h\|_{L^{p_3}}, \quad \frac{1}{p_1} + \frac{1}{p_2} + \frac{1}{p_3} = 1.$$

The embedding $H^1 \hookrightarrow L^p$ for $p \leqslant 6$ holds in two and three dimensions and we get

$$\begin{aligned} &\nu\|\nabla(Q_m(\mathbf{v}) - Q_m(\mathbf{w}))\|^2 \\ &\qquad \leqslant \|\mathbf{v}-\mathbf{w}\|_{L^3}\,\|\nabla Q_m(\mathbf{v})\|\,\|Q_m(\mathbf{v}) - Q_m(\mathbf{w})\|_{L^6} \\ &\qquad \leqslant \|\nabla(\mathbf{v}-\mathbf{w})\|\,\|\nabla Q_m(\mathbf{v})\|\|\nabla(Q_m(\mathbf{v}) - Q_m(\mathbf{w}))\|. \end{aligned}$$

Hence by $\|\nabla Q_m(\mathbf{v})\| \leqslant R$

$$\|\nabla(Q_m(\mathbf{v}) - Q_m(\mathbf{w}))\| \leqslant \nu^{-1}R\|\nabla(\mathbf{v}-\mathbf{w})\|.$$

This shows that $Q_m : \mathcal{V}_m \to \mathcal{V}_m$ is Lipschitz continuous and therefore continuous. Brouwer's fixed-point theorem guarantees the existence of a solution $v_m \in \mathcal{V}_m$ to the finite dimensional problem (2.17) for every $m \geqslant 1$.

(*v.3*) *Convergence* $m \to \infty$*:* the solutions $v_m \in \mathcal{V}_m$ form a bounded sequence in $\mathcal{V}_0$. As the embedding $H^1 \hookrightarrow L^2$ is continuous, there exists a subsequence $(\mathbf{v}_{m'}) \in \mathcal{V}_0$ that weakly converges in $\mathcal{V}_0$ and strongly in L^2 to a limit $\mathbf{v} \in \mathcal{V}_0$:

$$(\nabla(\mathbf{v}'_m - \mathbf{v}), \nabla\phi) \xrightarrow[m'\to\infty]{} 0 \quad \forall \phi \in \mathcal{V}_0, \quad \|\mathbf{v}_{m'} - \mathbf{v}\| \xrightarrow[m'\to\infty]{} 0.$$

This limit $\mathbf{v} \in \mathcal{V}_0$ solves (2.16), since for a sequence $\phi'_m \to \phi \in \mathcal{V}_0$ (strongly) it holds:

$$\begin{aligned}(\nabla \mathbf{v}'_m, \nabla \phi_m) &\to (\nabla \mathbf{v}, \nabla \phi) \qquad m' \to \infty \\ ((\mathbf{v}_{m'} \cdot \nabla)\mathbf{v}_m, \phi) &\to ((\mathbf{v} \cdot \nabla)\mathbf{v}, \phi) \quad m' \to \infty.\end{aligned}$$

$\phi \in \mathcal{V}_0$ is arbitrary, hence $\mathbf{v} \in \mathcal{V}_0$ is a solution of the incompressible Navier-Stokes equations in the space of divergence free functions:

$$(\nabla \mathbf{v}, \nabla \phi) + ((\mathbf{v} \cdot \nabla)\mathbf{v}, \phi) = (\mathbf{f}, \phi) \quad \forall \phi \in \mathcal{V}_0.$$

(*v.4*) *Uniqueness of the velocity:* Let $\mathbf{v}_1, \mathbf{v}_2 \in \mathcal{V}_0$ be two solutions to (2.16), such that

$$\|\nabla \mathbf{v}_i\| \leqslant \nu^{-1} \|f\|_{-1}, \quad i = 1, 2.$$

For $w := \mathbf{v}_1 - \mathbf{v}_2 \in \mathcal{V}_0$ it holds:

$$\nu \|\nabla \mathbf{w}\| = ((\mathbf{v}_1 \cdot \nabla)\mathbf{v}_1 - (\mathbf{v}_2 \cdot \nabla)\mathbf{v}_2, \mathbf{w}) = ((\mathbf{w} \cdot \nabla)\mathbf{v}_2, \mathbf{w}) + ((\mathbf{v}_1 \cdot \nabla)\mathbf{w}, \mathbf{w}).$$

Using (2.14) and Hölder's inequality

$$\nu \|\nabla \mathbf{w}\|^2 = (\mathbf{w} \cdot \nabla \mathbf{v}_2, \mathbf{w}) \leqslant \|\mathbf{w}\|_{L^3} \|\nabla \mathbf{v}_2\| \|\mathbf{w}\|_{L^6} \leqslant c^2 \|\nabla \mathbf{w}\|^2 \nu^{-1} \|\mathbf{f}\|_{-1},$$

and hence,

$$\|\nabla \mathbf{w}\|^2 \big(1 - c^2 \nu^{-2} \|\mathbf{f}\|_{-1}\big) \leqslant 0.$$

If $c^2 \nu^{-2} \|\mathbf{f}\|_{-1} \leqslant 1$ it follows that $\mathbf{w} = 0$.

(*p*) *existence of the pressure:* Given $\mathbf{v} \in \mathcal{V}_0 \subset \mathcal{V}$ the pressure $p \in \mathcal{L}$ must satisfy the equation

$$(p, \nabla \cdot \phi) = \underbrace{(\mathbf{f}, \phi) - \nu(\nabla \mathbf{v}, \nabla \phi) - ((\mathbf{v} \cdot \nabla)\mathbf{v}, \phi)}_{=: l(\phi)} \quad \forall \phi \in \mathcal{V}.$$

The right hand side $l : \mathcal{V} \to \mathbb{R}$ defines a linear functional and the existence of a unique solution follows as in the linear Stokes case.

Finally, the a priori bound follows by using the inf-sup inequality

$$\begin{aligned}\gamma \|p\| &\leqslant \sup_{\phi \in \mathcal{V}} \frac{(p, \nabla \cdot \phi)}{\|\nabla \phi\|} = \sup_{\phi \in \mathcal{V}} \frac{(\mathbf{f}, \phi) - \nu(\nabla \mathbf{v}, \nabla \phi) - ((\mathbf{v} \cdot \nabla)\mathbf{v}, \phi)}{\|\nabla \phi\|} \\ &\leqslant \|\mathbf{f}\|_{-1} + \nu \|\nabla \mathbf{v}\| + \|\mathbf{v}\|_{L^3} \|\nabla \mathbf{v}\|.\end{aligned}$$

□

The incompressible Navier-Stokes problem with homogenous Dirichlet values has a solution $\{\mathbf{v}, p\} \in \mathcal{V} \times \mathcal{L}$ for all Reynolds numbers and all right hand sides $\mathbf{f} \in H^{-1}(\Omega)$. This solution is unique only if the Reynolds number is very small:

$$Re \leqslant \sqrt{\frac{1}{c^2 \|\mathbf{f}\|_{-1}}}.$$

Most application problems however deal with high Reynolds numbers $Re \gg 1\,000$ and a unique solution cannot be guaranteed. As we know that flows at very high Reynolds numbers get turbulent, we cannot expect a unique result for arbitrary Reynolds numbers. The gap between theory and observation however is still very large.

Nearly no theoretical results are known for different boundary conditions, in particular for outflow conditions like the *do-nothing* condition. Here, it is even unknown, whether the homogenous problem

$$-\frac{1}{Re}\Delta\mathbf{v} + (\mathbf{v}\cdot\nabla)\mathbf{v} + \nabla p = 0, \quad \nabla\cdot\mathbf{v} = 0,$$

with homogenous boundary conditions

$$\mathbf{v} = 0 \text{ on } \Gamma^D, \quad \frac{1}{Re}\partial_n\mathbf{v} - p\mathbf{n} = 0 \text{ on } \Gamma^{out}$$

only has the trivial solution $\mathbf{v} = 0$ and $p = 0$ or if other non-trivial solutions exist.

Finally, we cite a regularity result for the stationary Navier-Stokes equations which is in agreement to the expectation:

Lemma 2.36 (Regularity of the Navier-Stokes solution)**.** Let $\Omega \subset \mathbb{R}^d$ be a convex polygonal or smooth domain of class $C^{2,1}$. Further, let $\bar{\mathbf{v}}^D \in H^2(\Omega)^d$ be a smooth extension of the Dirichlet data $\mathbf{v}^D$ on $\partial\Omega$ into the domain. Finally, let $\mathbf{f} \in L^2(\Omega)^d$. The solution to the Navier-Stokes equations has the regularity $\mathbf{v} \in H^2(\Omega) \cap \mathcal{V}$ and $p \in H^1(\Omega) \cap \mathcal{L}$ and it holds

$$\|\nabla^2\mathbf{v}\| + \|\nabla p\| \leqslant c_s\{\|\mathbf{f}\| + \|\nabla^2\bar{\mathbf{v}}^D\|\},$$

where the stability constant is related to the Reynolds number $c_s \sim Re$.

Next, let Ω be a C^{k+2}-domain and $\mathbf{f} \in H^k(\Omega)^d$. Then, every solution $\mathbf{v} \in H_0^1(\Omega)^d$ and $p \in L^2(\Omega)$ of the stationary Navier-Stokes equations has the regularity

$$\|\mathbf{v}\|_{H^{k+2}(\Omega)} + \|p\|_{H^{k+1}(\Omega)} \leqslant c\|\mathbf{f}\|_{H^k(\Omega)}.$$

Proof. For a proof of this result we refer to the literature, see Girault and Raviart [15] or Sohr [37]. □

2.2.2 The non-stationary Navier-Stokes equations

Finally, we discuss the non-stationary Navier-Stokes equations

$$\begin{aligned} \mathbf{v} &= \mathbf{v}^{\text{in}} && t = 0, \\ (\partial_t \mathbf{v}, \phi) + ((\mathbf{v}\cdot\nabla)\mathbf{v}, \phi) + \nu(\nabla\mathbf{v}, \nabla\phi) - (p, \nabla\cdot\phi) &= (\mathbf{f}, \phi) && \forall \phi \in \mathcal{V}, \\ (\nabla\cdot\mathbf{v}, \xi) &= 0 && \forall \xi \in \mathcal{L}. \end{aligned}$$

Like in the stationary case, we can restrict the problem to the space of divergence free functions $\mathcal{V}_0 \subset \mathcal{V}$. Integration of the variational formulation over the time-interval $I = [0, T]$ gives

$$\int_I \{(\partial_t \mathbf{v}, \phi) + ((\mathbf{v}\cdot\nabla)\mathbf{v}, \phi) + \nu(\nabla\mathbf{v}, \nabla\phi)\}\, dt = \int_I (\mathbf{f}, \phi)\, dt.$$

To analyze this variational formulation, we must first specify suitable function spaces. For the velocity part, natural choices for $\mathbf{v}$ and test function ϕ are

$$\mathbf{v}, \phi \in L^2(I; \mathcal{V}_0),$$

the space of square-integrable functions in time that map into $\mathcal{V}_0$. For the time-derivative of the velocity, we further ask for

$$\partial_t \mathbf{v} \in L^2(I; H^{-1}(\Omega)).$$

We denote this space by $W(0, T)$

$$W(0,T) := \{\phi \in L^2(I; \mathcal{V}_0),\ \partial_t \phi \in L^2(I; H^{-1}(\Omega))\}. \tag{2.20}$$

The spaces

$$\mathcal{V}_0 \subset H_0^1(\Omega)^d \subset L^2(\Omega)^d \cong [L^2(\Omega)^d]^* \subset H^{-1}(\Omega)$$

constitute a Gelfand triple and it holds (see [38])

$$W(0,T) \hookrightarrow C(\bar{I}; L^2(\Omega)^d).$$

Every function $\mathbf{v} \in W(0, T)$ is almost everywhere equal to a continuous function in time that maps into $L^2(\Omega)^d$. It remains to discuss the nonlinearity: does for functions $\mathbf{v}, \phi \in W(0, T)$ hold that

$$\int_I ((\mathbf{v}\cdot\nabla)\mathbf{v}, \phi)\, dt < \infty?$$

An answer is given by the following result:

Lemma 2.37. Let $\Omega \subset \mathbb{R}^d$ be an open set. For $d = 2$ it holds

$$\|\mathbf{v}\|_{L^4(\Omega)} \leqslant c\|\mathbf{v}\|^{\frac{1}{2}} \|\nabla\mathbf{v}\|^{\frac{1}{2}}.$$

In the case $d = 3$ it holds

$$\|\mathbf{v}\|_{L^4(\Omega)} \leqslant c\|\mathbf{v}\|^{\frac{1}{2}} \|\nabla\mathbf{v}\|^{\frac{3}{2}}.$$

Proof. A proof is given by Temam [38]. □

We consider the two-dimensional case. By Hölder's inequality ($1 = \frac{1}{4} + \frac{1}{2} + \frac{1}{4}$) and this Lemma we get

$$((\mathbf{v} \cdot \nabla)\mathbf{v}, \phi) \leqslant c\|\mathbf{v}\|_{L^4}\|\nabla\mathbf{v}\| \, \|\phi\|_{L^4} \leqslant c\|\mathbf{v}\|^{\frac{1}{2}}\|\nabla\mathbf{v}\|^{\frac{3}{2}}\|\phi\|^{\frac{1}{2}}\|\nabla\phi\|^{\frac{1}{2}}.$$

Using the embedding $W(0,T) \hookrightarrow C(\bar{I}; L^2(\Omega))$ it follows for the temporal integral by using Hölder's inequality (in time)

$$\begin{aligned}\int_I ((\mathbf{v} \cdot \nabla)\mathbf{v}, \phi)\,dt &\\ &\leqslant c\|\phi\|^{\frac{1}{2}}_{C(\bar{I};L^2(\Omega))}\|\mathbf{v}\|^{\frac{1}{2}}_{C(\bar{I};L^2(\Omega))}\int_I \|\nabla\mathbf{v}\|^{\frac{3}{2}}\|\nabla\phi\|^{\frac{1}{2}}\,dt\\ &\leqslant c\|\phi\|^{\frac{1}{2}}_{W(0,T)}\|\mathbf{v}\|^{\frac{1}{2}}_{W(0,T)}\|\mathbf{v}\|^{\frac{3}{2}}_{W(0,T)}\|\phi\|^{\frac{1}{2}}_{W(0,T)}\\ &\leqslant c\|\mathbf{v}\|^2_{W(0,T)}\|\phi\|_{W(0,T)}.\end{aligned}$$

This is exactly the desired stability result for the variational formulation. The nonlinearity is not bound in the three-dimensional case, if we ask for $\mathbf{v}, \phi \in W(0,T)$. We cite the following results that can be found in Temam [38]:

Lemma 2.38 (Instationary Navier-Stokes equations)**.** Let $\Omega \subset \mathbb{R}^d$ be a Lipschitz domain and

$$\mathbf{f} \in L^2(I; H^{-1}(\Omega)), \quad \mathbf{v}^0 \in \mathcal{V}_0.$$

Then, the instationary Navier-Stokes equation has at least one solution for arbitrary Reynolds numbers. This solution is unique in the two dimensional case (for arbitrary Reynolds numbers) and it holds

$$\mathbf{v} \in L^2(I; \mathcal{V}_0), \quad \partial_t\mathbf{v} \in L^2(I; H^{-1}(\Omega)).$$

In the three-dimensional case, unity is usually not given, and the solution has the reduced regularity

$$\mathbf{v} \in L^{\frac{8}{3}}(I; L^4(\Omega)), \quad \partial_t\mathbf{v} \in L^{\frac{4}{3}}(I; H^{-1}(\Omega)).$$

It is remarkable that the non-stationary solution is unique for all Reynolds numbers, if we look at the two-dimensional problem. Working with the stationary equation, uniqueness is only guaranteed for small data assumptions.

To prove existence of global solutions, uniqueness and regularity of the three dimensional problem is one of the big open problems in applied mathematics, see [6].

3 Finite Elements for incompressible flows

This chapter discusses the numerical approximation of incompressible flows with finite elements. Focus is the proper treatment of the saddle point character (recall Remark 2.4). We will see that the most essential ingredient for numerics is a discrete version of the inf-sup condition (recall Theorem 2.28). We start with the linear Stokes problem. Velocity $\mathbf{v} \in \bar{\mathbf{v}}^D + \mathcal{V}$ and pressure $p \in \mathcal{L}$ are given in

$$\mathcal{V} = H_0^1(\Omega; \Gamma_D)^d, \quad \mathcal{L} := L^2(\Omega) \setminus \mathbb{R},$$

where $\Gamma_D \subset \partial\Omega$ is the part of the boundary where Dirichlet conditions for the velocity are prescribed. We denote by $\bar{\mathbf{v}}^D \in H^1(\Omega)^d$ an extension of the Dirichlet data into the domain. Mostly we just consider $\bar{\mathbf{v}}^D = 0$. Given $\Gamma_D \neq \partial\Omega$, i.e. if we have a free outflow or inflow boundary with the *do-nothing* condition we do not need to filter the pressure space and use $\mathcal{L} = L^2(\Omega)$. The stationary Stokes problem in variational formulation reads as

$$\begin{aligned} (\nabla \mathbf{v}, \nabla \phi) - (p, \nabla \cdot \phi) &= (\mathbf{f}, \phi) \quad \forall \phi \in \mathcal{V}, \\ (\nabla \cdot \mathbf{v}, \xi) &= 0 \qquad \forall \xi \in \mathcal{L}, \end{aligned} \tag{3.1}$$

recall Section 1.4.6 and Section 2.1.

Following the standard philosophy of finite elements as a Galerkin method we choose finite dimensional subspaces $V_h \subset \mathcal{V}$ and $L_h \subset \mathcal{L}$ and determine the velocity approximation $\mathbf{v}_h \in V_h$ and the pressure approximation $p_h \in L_h$ as solution to

$$\begin{aligned} (\nabla \mathbf{v}_h, \nabla \phi_h) - (p_h, \nabla \cdot \phi_h) &= (f, \phi_h) \quad \forall \phi_h \in V_h, \\ (\nabla \cdot \mathbf{v}_h, \xi_h) &= 0 \qquad \forall \xi_h \in L_h. \end{aligned} \tag{3.2}$$

We recall that (3.2) is a saddle point problem. For classifying different Galerkin approaches we define

Definition 3.1 (Conforming finite element pair)**.** A discretization $V_h \times L_h$ of the Stokes equations is called *conforming*, if

$$V_h \subset \mathcal{V} = H_0^1(\Omega; \Gamma_D)^d, \quad L_h \subset \mathcal{L} = L^2(\Omega).$$

Otherwise we call it *non-conforming*.

Conformity in the pressure space is easy to realize. In the contrary: it would be rather difficult to construct a meaningful discrete space that is not square-integrable. For the velocity space we can employ the following criterion (which is not sharp, there might be more conforming spaces).

Lemma 3.2 (H^m-conforming discrete spaces)**.** Let Ω be a bounded domain in $\mathbb{R}^d$ and let Ω_h be an admissible triangulation. For $m \geqslant 1$ let $V_h \subset C^{m-1}(\bar{\Omega})$ be a subspace of $m-1$ times continuously differentiable functions on Ω with $v_h|_T \in C^m(T), T \in \Omega_h, v_h \in V_h$. Then V_h is $H^m(\Omega)$-conforming.

Proof. On bounded domains continuity gives integrability. We consider the case $m = 1$. For $m > 1$, the result follows recursively by considering derivatives of order $m-1$.
Conformity with respect to H^1 is shown by using the definition of the weak derivative:

$$\phi \in C_0^\infty(\Omega): \quad -(\partial_i \phi, v_h)_\Omega = -\sum_{T\in\Omega_h} (\partial_i \phi, v_h)_T = \sum_{T\in\Omega_h} \Big\{ (\phi, \partial_i v_h)_T - \int_{\partial T} \mathbf{n}_i \phi v_h \, ds \Big\},$$

where $\mathbf{n}$ is the outward facing unit normal on the edge (or in 3d face) ∂T of the element T. Every edge appears either twice $e = \partial T_1 \cap \partial T_2$ or is part of the boundary $e \in \partial\Omega$. As v_h is continuous and ϕ is continuous and zero on the boundary it holds with $\mathbf{n}_1 = -\mathbf{n}_2$ (as seen from T_1 and T_2) that

$$-(\partial_i \phi, v_h)_\Omega = \sum_{T\in\Omega_h} (\phi, \partial_i v_h)_T =: (\phi, \partial_i^h v_h)_\Omega,$$

Usually we just write ∂_i instead of ∂_i^h. But we must have in mind that this derivative is defined within the elements $T \in \Omega_h$ only and not on the element boundaries.

This argument can be extended to higher degree derivatives to show H^m-conformity of C^{m-1}-functions. □

Now, let $V_h \times L_h$ be a finite element pair with a given basis

$$V_h = \operatorname{span}\{\phi_h^i,\ i = 1, \dots, N_h^v\}, \quad L_h = \operatorname{span}\{\xi_h^i,\ i = 1, \dots, N_h^p\}.$$

Then, every function $\mathbf{v}_h \in V_h$ and $p_h \in L_h$ is uniquely given as

$$\mathbf{v}_h = \sum_{i=1}^{N_h^v} \mathbf{v}_i \phi_h^i, \quad p_h = \sum_{i=1}^{N_h^p} \mathbf{p}_i \xi_h^i.$$

By $\mathbf{v}_i$ and $\mathbf{p}_i$ we denote scalar coefficients, by ξ_h^i scalar and by ϕ_h^i vector valued basis functions. The discrete Stokes equations (3.2) are equivalent to the following linear system of equations

$$\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ -\mathbf{B}^T & 0 \end{pmatrix} \begin{pmatrix} \mathbf{v} \\ \mathbf{p} \end{pmatrix} = \begin{pmatrix} \mathbf{b} \\ 0 \end{pmatrix},$$

with matrices $\mathbf{A}$ and $\mathbf{B}$ as well as the right hand side $\mathbf{b}$:

$$\mathbf{A} := (\nabla\phi_h^j, \nabla\phi_h^i)_{i,j=1}^{N_h^v, N_h^v}, \quad \mathbf{B} := -(\nabla\cdot\phi_h^j, \xi_h^i)_{i,j=1}^{N_h^p, N_h^v}, \quad \mathbf{b} := (f, \phi_h^i)_{i=1}^{N_h^v}$$

and $\mathbf{v} = (\mathbf{v}_1, \dots, \mathbf{v}_{N_h^v})^T$ and $\mathbf{p} = (\mathbf{p}_1, \dots, \mathbf{p}_{N_h^p})^T$ denoting the vectors containing the scalar coefficients. It holds

Lemma 3.3 (Matrix of the discrete Stokes problem)**.** Let $V_h \times L_h$ be a finite dimensional conforming Galerkin discretization. The matrix $\mathbf{A} \in \mathbb{R}^{N_h^v \times N_h^v}$ is symmetric positive definite while the complete Stokes matrix is positive semidefinite and anti-symmetric in the off-diagonal blocks.

Proof. Symmetry and positivity of $\mathbf{A}$ is obtained from the vector-Laplace operator. Further it holds

$$\left\langle \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ -\mathbf{B}^T & 0 \end{pmatrix} \begin{pmatrix} \mathbf{v} \\ \mathbf{p} \end{pmatrix}, \begin{pmatrix} \mathbf{v} \\ \mathbf{p} \end{pmatrix} \right\rangle = \langle \mathbf{A}\mathbf{v}, \mathbf{v}\rangle + \langle \mathbf{B}\mathbf{p}, \mathbf{v}\rangle - \langle \mathbf{B}^T\mathbf{v}, \mathbf{p}\rangle = \langle \mathbf{A}\mathbf{v}, \mathbf{v}\rangle \geqslant 0$$

which is zero for every $\mathbf{v} = 0$ and $\mathbf{p} \in \mathbb{R}^{N_h^p}$. □

Remark 3.4 (Symmetry and positivity)**.** One could get the idea of simplifying the system by multiplication of the divergence equation with -1 to get a symmetric matrix

$$\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}^T & 0 \end{pmatrix} \begin{pmatrix} \mathbf{v} \\ \mathbf{p} \end{pmatrix} = \begin{pmatrix} \mathbf{b} \\ 0 \end{pmatrix}.$$

Considering this matrix we would even loose semidefiniteness, as

$$\left\langle \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}^T & 0 \end{pmatrix} \begin{pmatrix} \mathbf{v} \\ \mathbf{p} \end{pmatrix}, \begin{pmatrix} \mathbf{v} \\ \mathbf{p} \end{pmatrix} \right\rangle = \langle \mathbf{A}\mathbf{v}, \mathbf{v}\rangle + \langle \mathbf{B}\mathbf{p}, \mathbf{v}\rangle + \langle \mathbf{B}^T\mathbf{v}, \mathbf{p}\rangle = \langle \mathbf{A}\mathbf{v}, \mathbf{v}\rangle + 2\langle \mathbf{B}\mathbf{p}, \mathbf{v}\rangle.$$

We cannot identify the sign of $\langle \mathbf{B}\mathbf{p}, \mathbf{v}\rangle$. △

3.1 Divergence free finite elements

To solve (3.2) it is a immediate idea to look for velocity spaces V_h that are strictly divergence free, i.e. it holds $\nabla\cdot\mathbf{v}_h = 0$ for $\mathbf{v}_h \in V_h$ or that are at least weakly divergence free with respect to the test space $\xi_h \in L_h$, i.e.

$$(\nabla\cdot\mathbf{v}_h, \xi_h) = 0 \qquad \forall \xi_h \in L_h.$$

Then, the solution to the Stokes equation is given by the reduced vector-Laplace problem: find $\mathbf{v}_h \in V_h$ such that

$$(\nabla\mathbf{v}_h, \nabla\phi_h) = (\mathbf{f}, \phi_h) \qquad \forall \phi_h \in V_h. \tag{3.3}$$

Given a conforming discretization $V_h \subset V \subset H_0^1(\Omega)^d$, unique solvability follows from the continuity and ellipticity of the scalar product $(\nabla\cdot, \nabla\cdot)$ on $V_h \times V_h$:

$$|(\nabla \mathbf{v}_h, \nabla \phi_h)| \leqslant \|\nabla \mathbf{v}_h\| \, \|\nabla \phi_h\|, \quad (\nabla \phi_h, \nabla \phi_h) = \|\nabla \phi_h\|^2 \geqslant c_p^{-2} \|\phi_h\|^2,$$

with Poincaré constant c_p. However, it turns out to be a difficult task to construct divergence free spaces V_h. A systematic approach based on simple function spaces (e.g. piecewise linears) is not known.

In two dimensions we can construct such divergence free functions by using the "scalar" rotation (curl)

$$\operatorname{rot} : \mathbb{R} \to \mathbb{R}^2, \quad \operatorname{rot} \Phi(x, y) := \begin{pmatrix} -\partial_y \Phi(x, y) \\ \partial_x \Phi(x, y) \end{pmatrix}.$$

Under the assumption that Φ is sufficiently smooth, it holds $\operatorname{div} \operatorname{rot} \Phi = 0$. For an arbitrary scalar space W_h we define

$$V_h^{\operatorname{div}} := \{\operatorname{rot} \psi, \ \psi \in W_h\}$$

as strictly divergence free space V_h^{div}. This space serves as approximation space in (3.3). The following questions need to be answered:

1. Is the space $V_h^{\operatorname{div}} \subset \mathcal{V}_0 \subset V = H_0^1(\Omega)^d$ H^1-conforming? Conformity requires continuity of V_h^{div} over the element edges, see Lemma 3.2. Functions in V_h^{div} are derivatives of functions in W_h. If $W_h \subset C^1(\bar{\Omega})$, then the space V_h^{div} is $V_h^{\operatorname{div}} \subset C^0(\bar{\Omega})$ such that V_h^{div} is a H^1-conforming finite element space.

2. How can we satisfy Dirichlet values in V_h^{div} on Γ^D? On the boundary, $V_h^{\operatorname{div}} \ni \mathbf{v}_h = \operatorname{rot} w_h$ must satisfy $0 = \mathbf{v}_h = \operatorname{rot} w_h$. This condition could be enforced by requiring $\nabla w_h = 0$ for all functions $w_h \in W_h$.

3. How can we construct a basis of V_h^{div}? It holds

   $$\dim(V_h^{\operatorname{div}}) = \dim(\operatorname{rot}(W_h)) \leqslant \dim(W_h).$$

   Let $\{\psi_1, \psi_2, \dots\}$ be a basis of W_h. Will $\{\operatorname{rot}(\psi_1), \operatorname{rot}(\psi_2), \dots\}$ be a basis of V_h^{div}?

4. What are the approximation properties of W_h? For $W_h = P^r$ (polynomials of degree r) it cannot hold $V_h^{\operatorname{div}} = [P^{r-1}]^2$ (as this space would not be divergence free).

Divergence free elements have significant advantages in terms of accuracy and conservation of physical principles like mass conservation (which then also holds for the discrete solution). The construction however is difficult. One example is based on the Argyris element that is an H^2-conforming element usable for the biharmonic equation $\Delta^2 u = f$, see Figure 3.1.

It can be shown, that the rotation rot on the Argyris element is a injection $\operatorname{rot} : W_h \to V_h^{\operatorname{div}}$. As a result, we obtain a basis of V_h^{div} by applying rot onto the basis of W_h.

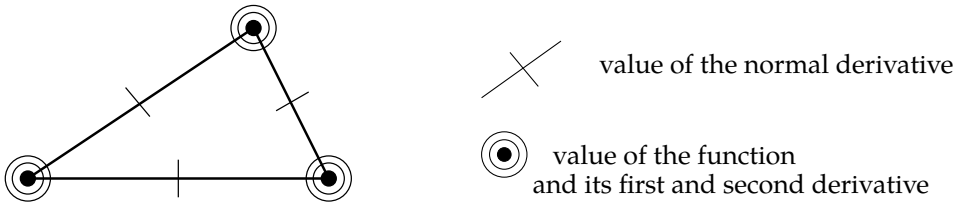


Figure 3.1: Nodal values of the Argyris element.

The Argyris element is a piecewise quintic element on triangles. Nodal values are prescribed in the corners of the triangles. Further, we prescribe first and second derivatives in the corners as well as the normal derivative on the edge midpoints. The divergence free functions in V_h^{div} are piecewise quartic functions.

Given a divergence free velocity $\mathbf{v}_h \in V_h^{div}$ of $v \in V$, the corresponding pressure $p_h \in L_h \subset L_0^2(\Omega)$ can be obtained as solution to

$$(p_h, \nabla \cdot \phi_h) = (\nabla \mathbf{v}_h, \nabla \phi_h) - (f, \phi_h) \quad \forall \phi_h \in V_h \subset H_0^1(\Omega)^d,$$

which corresponds to the linear system

$$-\mathbf{B}\mathbf{p} = \mathbf{A}\mathbf{v} - \mathbf{b}.$$

The unique solvability of this equation ($\mathbf{B} \in \mathbb{R}^{N_h^v \times N_h^p}$ with $N_h^v \neq N_h^p$) depends on the spaces L_h and V_h. Key is a discrete analogon to the *inf-sup condition*:

$$\min_{\xi_h \in L_h} \max_{\phi_h \in V_h} \frac{(\xi_h, \nabla \cdot \phi_h)}{\|\xi_h\| \, \|\nabla \phi_h\|} \geqslant \gamma_h \geqslant \gamma > 0.$$

We will discuss this inequality in detail. Considering piecewise quartic functions for V_h and piecewise quadratic functions for L_h the inf-sup condition will hold and the construction based on the Argyris element gives a divergence free space $V_h^{div} \subset V_h$ and an inf-sup stable pair $V_h \times L_h$ such that we find a pressure.

3.2 Stokes elements

We now consider the Stokes problem in saddle point form 3.2 and solve for velocity and pressure at the same time.

Let $V_h \times L_h \subset \mathcal{V} \times \mathcal{L}$ be a conforming finite element pair. The existence of a unique solution $\{\mathbf{v}_h, p_h\} \in V_h \times L_h$ is given by the following theorem.

Theorem 3.5 (Discrete Stokes problem)**.** Let $V_h \times L_h \subset V \times \mathcal{L}$ be a conforming finite element pair satisfying the discrete *inf-sup condition*

$$\min_{q_h \in L_h} \max_{\phi_h \in V_h} \frac{(q_h, \nabla \cdot \phi_h)}{\|q_h\| \, \|\nabla \phi_h\|} \geqslant \gamma_h \geqslant \gamma > 0.$$

Then, for every $\mathbf{f} \in H^{-1}(\Omega)^d$ there exists a unique solution $\{\mathbf{v}_h, p_h\} \in V_h \times L_h$ of the Stokes equation. It holds:

$$\|\nabla \mathbf{v}_h\| + \gamma_h \|p_h\| \leqslant c\|\mathbf{f}\|_{-1}.$$

Proof. (i) We start by defining a subspace of discretely divergence free functions $V_{h,0} \subset V_h$

$$V_{h,0} := \{\phi_h \in V_h | \ (\nabla \cdot \phi_h, \xi_h) = 0 \quad \forall \xi_h \in L_h\}. \tag{3.4}$$

It is not easy to estimate the dimension of the space $V_{h,0}$. It could even be that $\dim(V_{h,0}) = 0$ and the only possible solution would be $\mathbf{v}_h = 0$. We find $\mathbf{v}_h \in V_{h,0}$ as solution to the vector Laplace

$$(\nabla \mathbf{v}_h, \nabla \phi_h) = (\mathbf{f}, \phi_h) \quad \forall \phi_h \in V_{h,0}. \tag{3.5}$$

The existence of a unique solution $v_h \in V_{h,0}$ follows by linearity and ellipticity of $(\nabla\cdot, \nabla\cdot)$ in $V_{h,0} \subset H_0^1(\Omega)^d$. Further it holds

$$\|\nabla \mathbf{v}_h\|^2 = (f, \mathbf{v}_h) \leqslant \|\mathbf{f}\|_{-1}\|\mathbf{v}_h\|_1 \ \Rightarrow \ \|\nabla \mathbf{v}_h\| \leqslant c\|\mathbf{f}\|_{-1}.$$

(ii) Now, let $\mathbf{v}_h \in V_{h,0}$ be the solution from step (i). We find the pressure $p_h \in L_h$ as solution to

$$(p_h, \nabla \cdot \phi_h) = (\nabla \mathbf{v}_h, \nabla \phi_h) - (\mathbf{f}, \phi_h) \quad \forall \phi_h \in V_h. \tag{3.6}$$

Note that this problem is finite dimensional and equivalent to finding $\mathbf{p} \in \mathbb{R}^{N_h^p}$ such that

$$-\mathbf{B}\mathbf{p} = \mathbf{A}\mathbf{v} - \mathbf{b}$$

where

$$\mathbf{A} := \left(\nabla \phi_h^j, \nabla \phi_h^i\right)_{i,j=1}^{N_h^v, N_h^v}, \quad \mathbf{B} := -\left(\nabla \cdot \phi_h^j, \xi_h^i\right)_{i,j=1}^{N_h^p, N_h^v}, \quad \mathbf{b} := \left(\mathbf{f}, \phi_h^i\right)_{i=1}^{N_h^v}.$$

It holds $\mathrm{rg}(\mathbf{B}) = \ker(\mathbf{B}^T)^\perp$. This system has a solution if the right hand side $\mathbf{A}\mathbf{v} - \mathbf{b}$ is orthogonal on $\ker(-\mathbf{B}^T)$. This means

$$\langle \mathbf{A}\mathbf{v} - \mathbf{b}, \mathbf{z} \rangle = 0 \quad \forall \mathbf{z} \in \ker(-\mathbf{B}^T). \tag{3.7}$$

The *adjoint operator* $-\mathbf{B}^T$ is the discrete divergence, as

$$-\mathbf{B}^T \mathbf{z} = \left(\sum_{j=1}^{N_h^v} (\xi_h^i, \nabla \cdot \phi_h^j) \mathbf{z}_j \right)_{i=1}^{N_h^p} = (\nabla \cdot \mathbf{z}_h, \xi_h^i)_{i=1}^{N_h^p}.$$

Hence it holds $\ker(-\mathbf{B}^T) = V_{h,0}$ and with (3.5) we get

$$\mathbf{A}\mathbf{v} - \mathbf{b} \in \ker(-\mathbf{B}^T)^\perp$$

which corresponds to (3.7). There exists a pressure (at least one) $p_h \in L_h$ as solution to (3.6).

(*iii*) Let the discrete *inf-sup condition* hold. It is equivalent to the formulation

$$\max_{\phi_h \in V_h} \frac{(\xi_h, \nabla \cdot \phi_h)}{\|\nabla \phi_h\|} \geqslant \gamma_h \|\xi_h\| \quad \forall \xi_h \in L_h. \tag{3.8}$$

Let p_h^1, p_h^2 be two solutions to (3.6) for one velocity field $\mathbf{v}_h \in V_h$. For $q_h := p_h^1 - p_h^2$ it holds

$$-(q_h, \nabla \cdot \phi_h) = 0 \quad \forall \phi_h \in V_h,$$

and with (3.8) we get $\|q_h\| = 0$, since

$$0 = \max_{\phi_h \in V_h} \frac{(q_h, \nabla \cdot \phi_h)}{\|\nabla \phi_h\|} \geqslant \gamma_h \|q_h\|.$$

Further by (3.8)

$$\gamma_h \|p_h\| \leqslant \max_{\phi_h \in V_h} \frac{(p_h, \nabla \cdot \phi_h)}{\|\nabla \phi_h\|} = \max_{\phi_h \in V_h} \frac{(\nabla \mathbf{v}_h, \nabla \phi_h) - (\mathbf{f}, \phi_h)}{\|\nabla \phi_h\|} \leqslant \|\nabla \mathbf{v}_h\| + \|\mathbf{f}\|_{-1} \leqslant c\|\mathbf{f}\|_{-1}.$$

□

Remark 3.6. This theorem shows that every conforming finite element pair $V_h \times L_h$ admits a discrete solution $\{\mathbf{v}_h, p_h\}$. The velocity will always be unique and it can be bounded by the problem data. Uniqueness of the pressure and a corresponding estimate follows with help of the inf-sup condition. △

From this existence proof we identify the space $V_{h,0}$ as critical for the possible approximation property of a solution $\mathbf{v}_h \in V_{h,0} \subset V_h$, see (3.4) in the proof to Theorem 3.5. The space $V_{h,0}$ usually is no subspace of $\mathcal{V}_0$, as these functions are a.e. (in L^2) divergence free.

We cannot expect a best approximation property for the reduced velocity problem as it would hold for simple elliptic problems. Instead we must consider both the velocity and the pressure at the same time.

Lemma 3.7 (Stokes, best approximation)**.** Let $V_h \times L_h \subset \mathcal{V} \times \mathcal{L}$ be an inf-sup stable finite element approximation (3.2) of the Stokes problem (3.1). It holds:

$$\|\nabla(\mathbf{v} - \mathbf{v}_h)\| + \|p - p_h\| \leqslant c \left(\min_{\phi_h \in V_h} \|\nabla(\mathbf{v} - \phi_h)\| + \min_{\xi_h \in L_h} \|p - \xi_h\| \right),$$

where the constant $c > 0$ depends on the inf-sup constant γ_h. Further, on convex or smooth domains, it holds

$$\|\mathbf{v} - \mathbf{v}_h\| \leqslant ch \left(\min_{\phi_h \in V_h} \|\nabla(\mathbf{v} - \phi_h)\| + \min_{\xi_h \in L_h} \|p - \xi_h\| \right),$$

with constant $c = c(\gamma_h)$.

Proof. We define $\mathbf{e}_v := \mathbf{v} - \mathbf{v}_h \in \mathcal{V}$ and $e_p := p - p_h \in \mathcal{L}$. It holds by Galerkin orthogonality

$$\begin{aligned} (\nabla \mathbf{e}_v, \nabla \phi_h) &= (e_p, \nabla \cdot \phi_h) \quad \forall \phi_h \in V_h, \\ (\nabla \cdot \mathbf{e}_v, \xi_h) &= 0 \qquad\qquad\quad \forall \xi_h \in L_h. \end{aligned} \tag{3.9}$$

(i) First, we start with an estimate of the velocity error:

$$\|\nabla \mathbf{e}_v\|^2 = (\nabla \mathbf{e}_v, \nabla \mathbf{e}_v) - (e_p, \nabla \cdot \mathbf{e}_v) + (e_p, \nabla \cdot \mathbf{e}_v).$$

By Galerkin orthogonality, we get for arbitrary $\phi_h \in V_h$ and $\xi_h \in L_h$

$$\begin{aligned} \|\nabla \mathbf{e}_v\|^2 &= (\nabla \mathbf{e}_v, \nabla(\mathbf{v} - \phi_h)) - (e_p, \nabla \cdot (\mathbf{v} - \phi_h)) + (\nabla \cdot \mathbf{e}_v, p - \xi_h) \\ &\leqslant \|\nabla \mathbf{e}_v\| \, \|\nabla(\mathbf{v} - \phi_h)\| + \|e_p\| \, \|\nabla(\mathbf{v} - \phi_h)\| + \|\nabla \mathbf{e}_v\| \, \|p - \xi_h\|. \end{aligned}$$

Note that we have used Cauchy Schwarz and Lemma 2.6 in the previous inequality. By Young's inequality, we get for $\epsilon > 0$:

$$\|\nabla \mathbf{e}_v\| \leqslant (2 + \epsilon^{-1}) \|\nabla(\mathbf{v} - \phi_h)\| + 2\|p - \xi_h\| + \epsilon \|e_p\|. \tag{3.10}$$

(ii) Next, we estimate the pressure error. Let $\xi_h \in L_h$ be arbitrary

$$\|p - p_h\| \leqslant \|p - \xi_h\| + \|p_h - \xi_h\|. \tag{3.11}$$

For $p_h - \xi_h \in L_h$ we use the discrete inf-sup inequality to get

$$\begin{aligned} \gamma_h \|\xi_h - p_h\| &\leqslant \sup_{\phi_h \in V_h} \frac{(\xi_h - p_h, \nabla \cdot \phi_h)}{\|\nabla \phi_h\|} \\ &= \sup_{\phi_h \in V_h} \frac{(p - p_h, \nabla \cdot \phi_h)}{\|\nabla \phi_h\|} + \sup_{\phi_h \in V_h} \frac{(\xi_h - p, \nabla \cdot \phi_h)}{\|\nabla \phi_h\|} \end{aligned} \tag{3.12}$$

We use (3.9) on the first part to replace the pressure error e_p by the velocity error $\mathbf{e}_v$:

$$\sup_{\phi_h \in V_h} \frac{(e_p, \nabla \cdot \phi_h)}{\|\nabla \phi_h\|} = \sup_{\phi_h \in V_h} \frac{(\nabla \mathbf{e}_v, \nabla \phi_h)}{\|\nabla \phi_h\|} \leqslant \|\nabla \mathbf{e}_v\|.$$

Together with the second part of (3.12) we get the estimate

$$\gamma_h \|p_h - \xi_h\| \leqslant \|\nabla \mathbf{e}_v\| + \|p - \xi_h\|,$$

and finally, with (3.11) for $\|p - p_h\|$

$$\|e_p\| \leqslant (1 + \gamma_h^{-1}) \|p - \xi_h\| + \gamma_h^{-1} \|\nabla \mathbf{e}_v\|. \tag{3.13}$$

(iii) We insert this estimate into (3.10), using $\epsilon = \gamma_h/2$:

$$\|\nabla \mathbf{e}_v\| \leqslant c(\gamma_h) \left(\|\nabla(\mathbf{v} - \phi_h)\| + \|p - \xi_h\| \right).$$

Together with (3.13) we get the best-approximation property for the natural energy norm.

(*iv*) To derive the L^2-estimate we define the adjoint problem

$$(\nabla\phi, \nabla\mathbf{z}) - (\xi, \nabla\cdot\mathbf{z}) + (\nabla\cdot\phi, q) = \|\mathbf{e}_v\|^{-1}(\mathbf{e}_v, \phi) \quad \forall(\phi,\xi)\in\mathcal{V}\times\mathcal{L}.$$

As $\mathbf{e}_v/\|\mathbf{e}_v\| \in L^2$ it holds by Lemma 2.31 (if the domain has a convex or smooth boundary) that

$$\|\nabla^2\mathbf{z}\| + \|\nabla q\| \leqslant c_s \left\| \frac{\mathbf{e}_v}{\|\mathbf{e}_v\|} \right\| = c_s.$$

Now we choose the test functions $\phi := \mathbf{e}_v$ and $\xi := e_p$ and use the Galerkin orthogonality to insert the interpolants $I_h\mathbf{z} \in V_h$ and $I_h q \in L_h$. It follows:

$$\begin{aligned}
\|\mathbf{e}_v\| &= (\nabla\mathbf{e}_v, \nabla\mathbf{z}) - (e_p, \nabla\cdot\mathbf{z}) + (\nabla\cdot\mathbf{e}_v, q) \\
&= (\nabla\mathbf{e}_v, \nabla(\mathbf{z}-\phi_h)) - (e_p, \nabla\cdot(\mathbf{z}-\phi_h)) + (\nabla\cdot\mathbf{e}_v, (q-\xi_h)) \\
&\leqslant \|\nabla\mathbf{e}_v\|\,\|\nabla(\mathbf{z}-I_h\mathbf{z})\| + \|\nabla e_p\|\,\|\nabla(\mathbf{z}-I_h\mathbf{z})\| + \|\nabla\mathbf{e}_v\|\,\|q-I_h q\|
\end{aligned}$$

The result follows using the energy norm error estimate for $\|\nabla\mathbf{e}_v\| + \|e_p\|$ and the interpolation estimate:

$$\begin{aligned}
\|\mathbf{e}_v\| &\leqslant c_I h \left(\|\nabla\mathbf{e}_v\| + \|e_p\|\right)\left(\|\nabla^2\mathbf{z}\| + \|\nabla q\|\right) \\
&\leqslant c(\gamma_h)c_S h \left(\min_{\phi_h\in V_h} \|\nabla(\mathbf{v}-\phi_h)\| + \min_{\xi_h\in L_h} \|p-\xi_h\| \right).
\end{aligned}$$

□

The proof shows that velocity and pressure errors are not independently of each other. This result is expected since the divergence-free constraint of the velocity is only postulated weakly, tested with the pressure space. Let us give an alternative proof for Lemma 3.7. For this, we first show a discrete analogon to Lemma 2.30.

Lemma 3.8. Let $\mathcal{X}_h := V_h \times L_h$ be a conformal subspace of $V\times Q$ and let the discrete inf-sup condition hold true in $\mathcal{X}_h$. Then, for all $U_h \in V_h \times L_h$ the following estimation holds true

$$c|||U_h||| \leqslant \sup_{\Phi_h\in\mathcal{X}_h, |||\Phi_h|||=1} A(U_h, \Phi_h).$$

Proof. The proof follows analogously to Lemma 2.30. □

With this Lemma, we can now give an alternative proof for Lemma 3.7.

Proof. (*Alternative proof to Lemma 3.7*) (*i*) Let us use the notation $U := \{\mathbf{v}, p\} \in V\times Q$ and $U_h := \{\mathbf{v}_h, p_h\}$. We split the error $E_h = U - U_h$ into the interpolation error $\eta_h := U - i_h U$ and projection error $\chi_h := i_h U - U_h$. Then,

$$|||E_h||| \leqslant |||\eta_h||| + |||\chi_h|||.$$

For the interpolation error $|||\eta_h|||$ the estimation is given by the usual interpolation estimates. For the projection error, we use the previous Lemma 3.8:

$$c|||\chi_h||| \leqslant \sup_{\Phi_h \in \mathcal{X}_h, |||\Phi_h|||=1} A(\chi_h, \Phi_h).$$

Using $E_h = \eta_h + \chi_h$ and the Galerkin orthogonality $A(E_h, \Phi_h) = 0$ for all $\Phi_h \in \mathcal{X}_h$:

$$A(\chi_h, \Phi_h) = A(E_h, \Phi_h) - A(\eta_h, \Phi_h) = -A(\eta_h, \Phi_h).$$

Since the bilinear form A is continuous on $\mathcal{X}_h$ and $\mathcal{X}$ it holds

$$c|||\chi_h||| \leqslant c' \sup_{\Phi_h \in \mathcal{X}_h, |||\Phi_h|||=1} |||\eta_h||| |||\Phi_h||| = c'|||\eta_h|||.$$

Together it holds

$$|||E_h||| \leqslant c|||U - i_h U|||.$$

This means that the approximation error is bounded by the interpolation error.

(*ii*) For the L^2-error we consider the dual problem defined by:

$$A(\Phi, Z) = \|\mathbf{e}_v\|^{-1}(\mathbf{e}_v, \phi) \quad \forall \Phi = \{\phi, \xi\} \in \mathcal{X}.$$

Then, for $E_h := \{\mathbf{v} - \mathbf{v}_h, p - p_h\}$ it follows (using Galerkin orthogonality and continuity of the bilinear form):

$$\|\mathbf{v} - \mathbf{v}_h\| = A(E_h, Z) = A(E_h, Z - i_h Z) \leqslant c|||E_h||| \, |||Z - i_h Z|||.$$

The error estimate follows with the interpolation estimate for the dual problem and the a priori estimation for $\|\nabla(\mathbf{v} - \mathbf{v}_h)\| + \|p - p_h\|$. □

The approximation order of the Stokes element depends on the polynomial degree of the finite element pair:

Lemma 3.9 (Stokes, a priori estimate)**.** Let $V_h \times L_h$ be an inf-sup stable finite element pair of order k for the velocity and l for the pressure. Further, let $\mathbf{v} \in H^{k+1}(\Omega)^d$ and $p \in H^{l+1}(\Omega)$ be the solution to the incompressible Stokes equations. It holds

$$\|\nabla(\mathbf{v} - \mathbf{v}_h)\| + \|p - p_h\| \leqslant ch^{\min\{k,l+1\}} \Big(\|\nabla^{k+1}\mathbf{v}\| + \|\nabla^l p\| \Big). \tag{3.14}$$

Proof. First, by Lemma 3.7 it holds

$$\|\nabla(\mathbf{v} - \mathbf{v}_h)\| + \|p - p_h\| \leqslant c(\gamma_h^{-1}) \left(\|\nabla(\mathbf{v} - I_h^v \mathbf{v})\| + \|p - I_h^p p\| \right),$$

where $I_h^v \mathbf{v} \in V_h$ and $I_h^p p \in L_h$ are the nodal interpolations. Given sufficient regularity it holds with interpolation estimates

$$\|\nabla(\mathbf{v} - I_h \mathbf{v})\| \leqslant c_I h^k \|\nabla^{k+1}\mathbf{v}\|, \quad \|p - I_h p\| \leqslant c_I h^{l+1} \|\nabla^{l+1} p\|.$$

This completes the estimate. □

This lemma shows that the optimal degree for velocity and pressure space differs by one. If $l = k-1$, optimal order of convergence is given. Possible candidates for such finite element pairs are the Taylor-Hood element P^2-P^1 or Q^k-Q^{k-1} or the modified Taylor-Hood element with discontinuous pressure $Q^2-P^{1,dc}$. This element has the further advantage of local mass conservation. We will study these elements in more details in the following.

Remark 3.10 (Optimality of the a priori estimates)**.** In terms of mesh parameter $h > 0$, the estimates in Lemma 3.9 are optimal and represent the best-approximation property. They however exhibit two shortcomings which are severe under given circumstances.

First, only coupled estimates for velocity and pressure are given. Assume that the right hand side $\mathbf{f}$ is such that its divergence free part is zero with $\mathbf{f} = \nabla q$. Then, the Stokes equations have the unique solution $\mathbf{v} = 0$ and $p = q$. Equation (3.14) gives an estimate for the velocity error depending on the pressure error. And indeed, most standard approaches elements like Taylor-Hood or the Q^2-$P^{1,dc}$ element will show exactly this unsatisfactory behavior with very large errors. So called *gradient-robust mixed methods* are designed in such a way that the velocity approximation is independent of the pressure. See [25] for details. In most applications, the right hand side $\mathbf{f}$ itself is not critical, as it will be zero or a fixed gravity error. In large scale deformations however, Coriolis terms may have the same effect. In terms of fluid-structure interactions, the domain motion and the ALE map is a further source of such problems.

The second issue in Lemma 3.9 is the negative dependence of the error constant on the inf-sup constant. It is well known that the inf-sup constant depends on the shape of the domain and that it goes to zero for strongly anisotropic domains, see [11]. For very long channels, this would suggest large error constants. Here however, numerical reality is in favor, such that usual finite element approaches do not see this issue. The proof of Lemma 3.9 can be modified in such a way that Fortin's criteria (see later Lemma 3.13) is applied only locally, such that the bad behavior of the global inf-sup constant does impact the result. See [26, 27] for details. △

In the following, we investigate concrete choices for finite element pairs. We will study the discrete *inf-sup condition* and the approximation properties. We distinguish between conformal spaces $V_h \times L_h \subset V \times Q$ and non-conformal spaces. However, for the pressure we will always consider conformal spaces $L_h \subset L^2(\Omega)$, since even piecewise polynomial but not globally continuous functions are still in $L^2(\Omega)$.

Let us first introduce a triangulation Ω_h of our domain Ω into open triangles / tetrahedrons T or open rectangles / cuboids K. The triangulation shall fulfill the usual assumptions on structure and form and shape regularity. We denote

$$\mathcal{P}^r := \operatorname{span}\{x^i y^j,\ 0 \leqslant i+j \leqslant r\}, \quad \mathcal{Q}^r := \operatorname{span}\{x^i y^j,\ 0 \leqslant i,j \leqslant r\},$$

the polynomial spaces of degree r. It holds for $d = 2$ and $d = 3$:

$$\dim(\mathcal{P}^r)\Big|_{d=2} = \frac{(r+1)(r+2)}{2}, \quad \dim(\mathcal{P}^r)\Big|_{d=3} = 1+\frac{r^3+6r^2+11r+6}{6}, \quad \dim(\mathcal{Q}^r) = (r+1)^d$$

For the construction of the finite element ansatz we define the discrete spaces V_h und L_h using a parametric approach. We define by

$$\hat{K} := (0,1)^d, \quad \hat{T} := \left\{ x \in (0,1)^d,\ 0 < \sum_{i=1}^{d} x_i < 1 \right\},$$

a reference rectangle $\hat{Q}$ and a reference tetrahedron $\hat{T}$. Further, for each triangle $T \in \Omega_h$, and rectangle $K \in \Omega_h$, respectively, there exists a map $T_T : \hat{T} \to T$ and $T_K : \hat{K} \to K$, respectively. We assume that these mappings are sufficiently continuous differentiable with differentiable inverse mapping. Then, we define the spaces of continuous, piecewise polynomial functions as

$$P^r(\Omega) := \{\phi \in C(\bar{\Omega}) :\ \phi \circ T_T \in \mathcal{P}^r\}, \quad Q^r(\Omega) := \{\phi \in C(\bar{\Omega}) :\ \phi \circ T_K^{-1} \in \mathcal{Q}^r\}.$$

Further, the spaces of piecewise polynomial and globally discontinuous functions are given by

$$P^{r,dc}(\Omega) := \{\phi \in L^2(\Omega) :\ \phi \circ T_T^{-1} \in \mathcal{P}^r\}, \quad Q^{r,dc}(\Omega) := \{\phi \in L^2(\Omega) :\ \phi \circ T_K^{-1} \in \mathcal{Q}^r\}.$$

We call such finite element spaces *isoparametric*, if the transformation stems from the same space as the ansatz functions itself, i.e.

$$P^{r,iso}(\Omega) := \{\phi \in C(\bar{\Omega}) :\ \phi \circ T_T^{-1} \in \mathcal{P}^r, \quad T_T \in [\mathcal{P}^r]^d\}.$$

In general, we consider isoparametric spaces.

For preparing the analysis of different finite element pairs we cite two often used interpolation results.

Lemma 3.11 (Clement-Interpolation)**.** Let $u \in H^1(\Omega)$ and $V_h \subset H^1(\Omega)$ be a Lagrangian finite element space with basis $\phi_i(x_j) = \delta_{ij}$ on the triangulation Ω_h. Then, the Clement-Interpolation $C_h u \in V_h$ given as

$$C_h u = \sum_{i=1}^{n} \chi_i(u)\phi_i, \quad \chi_i(u) = \frac{1}{|P_i|}\int_{P_i} u dx,$$

with the *patches* P_i defined as unions of all elements that touch a node x_i

$$P_i := \bigcup_{K \in \Omega_h,\ x_i \in \bar{K}} K$$

is H^1-stable

$$\|\nabla C_h u\| \leqslant c\|\nabla u\| \quad \forall u \in H^1(\Omega).$$

It holds

$$\|\nabla C_h u\|_K \leqslant c_1 \|\nabla u\|_{P_K}, \quad P_K := \bigcup_{L \in \Omega_h,\ \bar{L} \cap \bar{K} \neq \emptyset} L.$$

Proof. The proof follows by showing stability of the node functionals $\xi_i(\cdot)$ and using Bramble-Hilbert Lemma. See [9]. □

Remark 3.12 (Interpolation on anisotropic meshes)**.** The Clement interpolation is an H^1-stable operator

$$\|\nabla C_h u\|_K \leqslant c\|\nabla u\|_{P_K},$$

with a constant $c > 0$ that does not depend on $h > 0$. The Clement operator however fails, if the mesh-elements $K \in \Omega_h$ are anisotropic with $h_{min}(K) \ll h_{max}(K)$. On such elements, it only holds

$$\|\nabla C_h u\|_K \leqslant c\frac{h_{max}(K)}{h_{min}(K)}\|\nabla u\|_{P_K}.$$

An H^1-stable alternative to the Clement operator, which is also stable on anisotropic meshes, is the Scott & Zhang operator. Here, the nodal values are also defined as averages, but averaging is only applied over edges of elements. This helps to avoid mixing of mesh-sizes in different directions. See [35] for basics on the Scott & Zhang interpolation operator, and [2] for an analysis of interpolation operators on anisotropic meshes. △

3.2.1 Conformal spaces with discontinuous pressure

In this section, we consider conformal ansatz spaces $V_h \times L_h \subset V \times Q$ with discontinuous pressure $L_h \subset L^2(\Omega)$. Since these pressure spaces contain the piecewise constant functions, they possess an important local conservation property: with the choice $\xi_h \in L_h$ with $\xi_h = 1$ on K and $\xi_h = 0$ for all $K' \neq K$ it follows from the divergence-free constraint

$$0 = (\nabla \cdot \mathbf{v}_h, \xi_h) = \int_K \nabla \cdot \mathbf{v}_h \, dx = \int_{\partial K} n \cdot \mathbf{v}_h \, do,$$

the *local mass conservation*. In application problems, local conservation properties are as important as global approximation properties. We summarize the different elements in 3.2 (left panel).

a) The $P^1 - P^{0,\mathbf{dc}}$ and $Q^1 - P^{0,\mathbf{dc}}$-elements. These elements are the easiest conform triangle and rectangle element elements for the Stokes equations. Both elements are not suitable.
The space of the weakly divergence free functions $V_{h,0}$ contains in the case of the $P^1 - P^{0,dc}$-element almost only the zero. This can be seen in the following construction: we consider the excerpt form a regular grid with grid size h as shown in Figure 3.3. At the boundary let $\mathbf{v}_h$ have homogeneous Dirichlet conditions. For the coefficients in the Galerkin ansatz $\mathbf{v}_h = \sum_{i=1}^{N_h^v} \mathbf{v}_i \phi_h^i$, let us use the notation $\mathbf{v} = (\mathbf{v}_i)_{i=1}^{N_h^v} = (\{\mathbf{v}_i^x, \mathbf{v}_i^y\})_{i=1}^{N_h^v}$. Let us now determine the value $\mathbf{v}_i$ in the node point $x \in \Omega_h$ (see Figure 3.3). For this, let ϕ be the nodal basis

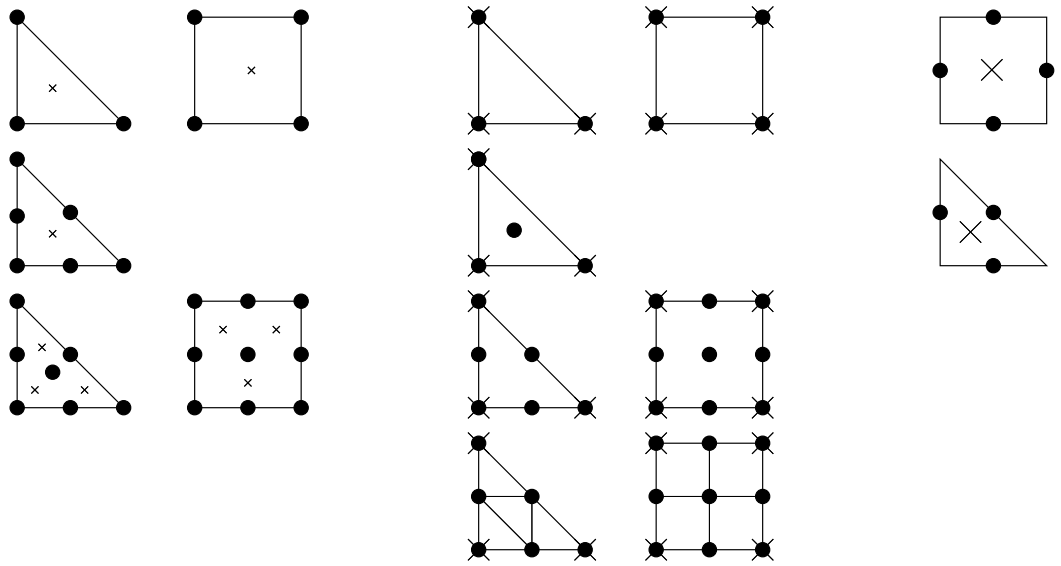


Figure 3.2: Conformal Stokes elements. Velocity degrees of freedom are denoted by a dot, pressure degrees of freedom are marked by a cross. Left: elements with discontinuous pressure: non stable $P^1 - P^{0,dc}$ and $Q^1 - P^{0,dc}$ elements as well as stable $P^2 - P^{0,dc}$, and the enriched $P^{2,b} - P^{1,dc}$ bulb-element and the $Q^2 - P^{1,dc}$ element. Middle: elements with continuous pressure: the non stable *equal-order* elements $P^1 - P^1$ and $Q^1 - Q^1$, the mini-element $P^{1,b} - P^1$, the Taylor-Hood elements and the *iso*-elements. Right: the stable but non-conformal $P^{1,nc} - P^{0,dc}$ and the rotated bilinear element $Q^{1,rot} - P^{0,dc}$ (Rannacher-Turek element).

function in P^1 which is 1 at x and 0 in all other points. On each triangle element K_i it holds that $\nabla\phi|_{K_i}$ is constant. It holds (compare Figure 3.3) for the triangles K_1 and K_2:

$$\nabla\phi\Big|_{K_1} = h^{-1}\begin{pmatrix}1\\1\end{pmatrix}, \quad \nabla\phi\Big|_{K_2} = h^{-1}\begin{pmatrix}0\\1\end{pmatrix},$$

From the local divergence-free property together with $\xi_h|_{K_j} = 1$ and $\xi_h = 0$ else it follows:

$$0 = \int_{K_j} \nabla\cdot\mathbf{v}_h\,dx = \int_{K_j} \mathbf{v}_i^x\partial_x\phi + \mathbf{v}_i^y\partial_y\phi\,dx = |K_j|\Big(\mathbf{v}_i^x\partial_x\phi + \mathbf{v}_i^y\partial_y\phi\Big)\Big|_{K_j}.$$

Therefore, it holds:

$$\mathbf{v}_i^x + \mathbf{v}_i^y = 0, \quad \mathbf{v}_i^y = 0 \quad \Rightarrow \quad \mathbf{v}_i = 0.$$

This construction can be repeated for all other triangles. Therefore, it holds

$$(\nabla\cdot\mathbf{v}_h, \xi_h) = 0 \quad \forall\xi_h \in P^{0,dc} \quad \Rightarrow \quad \mathbf{v}_h = 0.$$

The $Q^1 - P^{0,dc}$-element produces a larger space $V_{h,0}$. However, the *inf-sup condition* is not fulfilled and therefore the element is not stable. In order to see this, let us consider the kernel

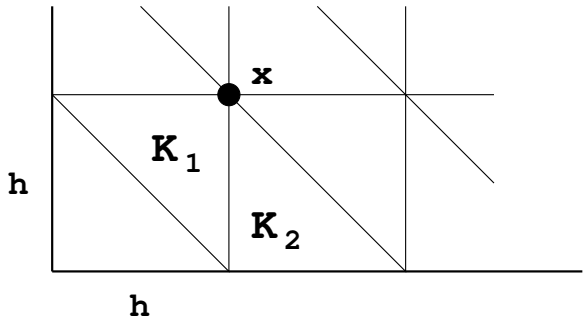


Figure 3.3: Computing the space $V_{h,0}$ for the $P^1 - P^{0,dc}$-element.

of the operator grad : $L_h \to V_h$. It is characterized by:

$$\mathrm{Kern}(\mathrm{grad}) = \{\xi_h \in P^{0,dc}(\Omega) : \quad (\xi_h, \nabla \cdot \phi_h) = 0 \quad \forall \phi_h \in V_h\}.$$

On a uniform mesh with squares of mesh size h, we choose for an inner nodal point $x_i \in \Omega_h$ the test function (see Figure 3.4)

$$\psi := \begin{pmatrix} \alpha \\ \beta \end{pmatrix} \phi_i(x, y).$$

Then, for piecewise constant $\xi_K := \xi_{h|K}$ it holds:

$$(\xi_h, \nabla \cdot \psi) = \sum_{K \in \Omega_h} \xi_K \int_K \nabla \cdot \psi \, dx = \sum_{K \in \Omega_h} \xi_K \int_{\partial K} n \cdot \psi \, do.$$

Outside of a *patch* of four elements around the nodal point x_i, the function is $\psi \equiv 0$. With the notation of Figure 3.4 it holds for arbitrary α and β:

$$\begin{aligned} 0 &= \frac{\xi_1}{h}(\alpha + \beta) + \frac{\xi_2}{h}(-\alpha + \beta) + \frac{\xi_3}{h}(-\alpha - \beta) + \frac{\xi_4}{h}(\alpha - \beta) \\ &= \frac{\alpha}{h}(\xi_1 - \xi_2 - \xi_3 + \xi_4) + \frac{\beta}{h}(\xi_1 + \xi_2 - \xi_3 - \xi_4) \end{aligned}$$

From this, it follows $\xi_1 = \xi_3$ and $\xi_2 = \xi_4$. With the additional normalization condition of the pressure, we obtain:

$$0 = \int_\Omega \xi_h \, dx = \sum_{K \in \Omega_h} \xi_K$$

Therefore, it is $\xi_1 = \xi_3 = -\xi_2 = -\xi_4 = \xi_c$ with an arbitrary constant $\xi_c \in \mathbb{R}$. Pressure functions with this alternating pattern $q_h = \pm 1$ lie in the kernel of the operator grad. Thus, the Stokes equation cannot be solved uniquely and the *inf-sup Bedingung* does not hold. The resulting pressure pattern from Figure 3.4 is called *checkerboard-pattern*.

If we restrict the pressure space L_h to the complement of the kernel of the operator grad:

$$\tilde{L}_h := \{p_h \in L_h : \ (p_h, q_h) = 0 : \ \forall q_h \in \mathrm{Kern}(\mathrm{grad})\},$$

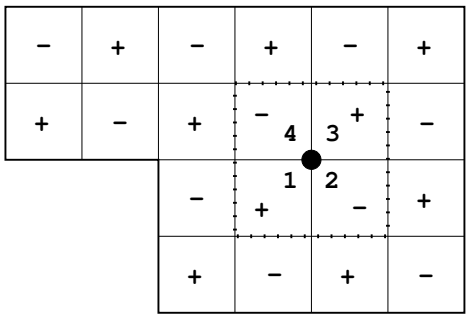


Figure 3.4: Checkerboard instability. Patch around the node point x_i.

then we can show an *inf-sup condition* of type

$$\min_{\xi_h \in L_h} \max_{\phi_h \in V_h} \frac{(\xi_h, \nabla \cdot \phi_h)}{\|\xi_h\| \, \|\nabla \phi_h\|} \geqslant \gamma h$$

in this smaller space. The factor h cannot be avoided, such that there is no sufficient (grid independent) stability. Although the element $Q^1 - P^{0,dc}$ is not *inf-sup* stable, it is still used quite often. For the velocity, we have convergence of first order.

b) Quadratic velocities with discontinuous pressures. The preceding element did not have enough degrees of freedom in the velocity space (compared to the pressure space) to give stability.

For the following we give a flexible criterion for showing inf-sup stability, see [5].

Lemma 3.13 (Fortin criterion)**.** Let $V_h \times L_h \subset \mathcal{V} \times \mathcal{L}$ be a finite element pair. Given a H^1-stable projection operator $\pi_h : \mathcal{V} \to V_h$ satisfying

$$\|\nabla \pi_h \phi\| \leqslant c_\pi \|\nabla \phi\| \quad \forall \phi \in \mathcal{V}, \quad (\nabla \cdot (\phi - \pi_h \phi), \xi_h) = 0 \quad \forall \xi_h \in L_h,$$

it holds

$$\inf_{\xi_h \in L_h} \sup_{\phi_h \in V_h} \frac{(\xi_h, \nabla \cdot \phi_h)}{\|\nabla \phi_h\| \, \|\xi_h\|} \geqslant \gamma_h := \gamma c_\pi^{-1},$$

where $\gamma > 0$ is the continuous inf-sup constant in $\mathcal{V} \times \mathcal{L}$.

Proof. Let $p_h \in L_h \subset \mathcal{L}$. It holds with the continuous inf-sup condition

$$\gamma \|p_h\| \leqslant \sup_{\phi \in \mathcal{V}} \frac{(p_h, \nabla \cdot \phi)}{\|\nabla \phi\|} = \sup_{\phi \in \mathcal{V}} \frac{(p_h, \nabla \cdot (\phi - \pi_h \phi))}{\|\nabla \phi\|} + \sup_{\phi \in \mathcal{V}} \frac{(p_h, \nabla \cdot \pi_h \phi)}{\|\nabla \phi\|}.$$

As the first part is zero due to the orthogonality of the projection π_h it further follows with the stability of the projection

$$\gamma \|p_h\| \leqslant \sup_{\phi \in \mathcal{V}} \frac{(p_h, \nabla \cdot \pi_h \phi)}{\|\nabla \pi_h \phi\|} \sup_{\phi \in \mathcal{V}} \frac{\|\nabla \pi_h \phi\|}{\|\nabla \phi\|} \leqslant c_\pi \sup_{\phi_h \in V_h} \frac{(p_h, \nabla \cdot \phi_h)}{\|\nabla \phi_h\|},$$

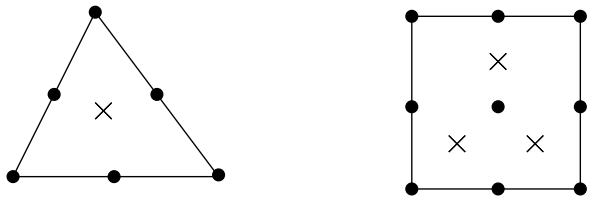


Figure 3.5: Modified Taylor-Hood elements $P^2 - P^{0,dc}$ (left) and $Q^2 - P^{1,dc}$ (right). Circles denote (continuous) nodal values of the basis functions, crosses stand for monomial values of a discontinuous approach.

as $\pi_h \phi \in V_h$. □

For some elements the Fortin criterion Lemma 3.13 helps to show inf-sup stability:

Lemma 3.14 (Modified Taylor-Hood elements with discontinuous pressure)**.** The $P^2 - P^{0,dc}$ and $Q^2 - P^{1,dc}$ elements are inf-sup stable.

Proof. See Figure 3.5 for a sketch of these two element pairs. We construct a projection operator $\pi_h : \mathcal{V} \to V_h$ that has both properties, H^1-stability and the required orthogonality.

(*i*) *The triangular element.* We construct π_h as $\pi_h := C_h + E_h$, where $C_h : \mathcal{V} \to V_h^1$ is the Clement operator from Lemma 3.11 interpolating to the space of piecewise linear functions. This space has three degrees of freedom (for every velocity component) and fixes the three nodal points of a triangle. This operator C_h satisfies

$$\|\nabla C_h \mathbf{v}\|_K \leqslant c \|\nabla \mathbf{v}\|_{P(K)},$$

where $P(K)$ is a patch of elements around K. See Lemma 3.11 for details. It remains to fulfill the orthogonality condition. As the pressure space is discontinuous, it holds on every K choosing by $\xi_h \equiv 1$ on K and $\xi_h = 0$ elsewhere:

$$(\nabla \cdot (\mathbf{v} - \pi_h \mathbf{v}), \xi_h) = \int_K \nabla \cdot (\mathbf{v} - \pi_h \mathbf{v}) \, dx = \int_{\partial K} \mathbf{n} \cdot (\mathbf{v} - \pi_h \mathbf{v}) \, do.$$

For $\pi_h := C_h + E_h$ one condition is imposed on every edge $e \in \partial K$:

$$\int_e \mathbf{n} \cdot E_h \mathbf{v} \, do = \int_e \mathbf{n} \cdot (\mathbf{v} - C_h \mathbf{v}) \, do.$$

This is easily established by the remaining degrees of freedom (two per edge).

(*ii*) *The quadrilateral element.* We define the projection as $\pi_h := C_h + E_h + B_h$, where C_h again is a H^1-stable Clement interpolation, E_h takes care of the edges and B_h of the additional middle degree of freedom. For the orthogonality it holds for $\xi_K \in L_h$ with $\xi_K = 0$ for all $K' \neq K$:

$$(\nabla \cdot (\mathbf{v} - \pi_h \mathbf{v}), \xi_h)_K = -(\mathbf{v} - \pi_h \mathbf{v}, \nabla \xi_h)_K + \int_{\partial K} \mathbf{n} \cdot (\mathbf{v} - \pi_h \mathbf{v}) \xi_h \, do.$$

As ξ_h is piecewise linear, $\nabla\xi_h \in \mathbb{R}^2$ is a constant vector on every element. The two inner degrees of freedom are used to define the operator B_h via

$$\int_K B_h \mathbf{v}_i \, dx = \int_K \mathbf{v}_i - \pi_h \mathbf{v}_i \, dx, \quad i = 1, 2.$$

Finally, $\xi_h \in L_h$ is a linear function on every edge $e \in \partial K$, and the two remaining degrees of freedom are required for satisfying

$$\int_e (\mathbf{n} \cdot E_h \mathbf{v}) \xi_h \, do = \int_e \mathbf{n} \cdot (\mathbf{v} - \pi_h \mathbf{v}) \xi_h \, do.$$

□

The $Q^2 - P^{1,dc}$ element is an excellent mixed finite element for the discretization of incompressible flows. Quadratic velocities are a good compromise between high accuracy at acceptable computational effort (as the effort is increasing in powers of the polynomial degree). Discontinuous pressures give local conservation. Finally, in the context of fluid-structure interactions, discontinuous pressures simplify the coupling to a possibly incompressible solid that also has a pressure variable (the coupling between two pressures at the interface is discontinuous). See [40, 41] for applications.

The $P^2 - P^{0,dc}$ element is not order optimal. Despite the quadratic velocities we obtain because of the low order in the pressure only

$$\|\nabla(\mathbf{v} - \mathbf{v}_h)\| + \|p - p_h\| = O(h).$$

c) The enriched bulb element $P^{2,b} - P^{1,\mathbf{dc}}$. The $P^2 - P^{1,dc}$-element is not stable. In order to achieve stability, we need to enrich the velocity space: in each triangle K we define the *bulb*-function (here in the reference triangle $\hat{K}$)

$$b_K(x, y) = xy(h - x - y),$$

which vanishes on the boundary of the triangle and we add inner degrees of freedom:

$$V_h := \left[P^2 + \text{span}\{\beta_K b_K, \ K \in \Omega_K\}\right]^2.$$

The pressure is piecewise linear, i.e. $\nabla\xi_h$ is a constant for each element $K \in \Omega_h$. The orthogonality condition for π_h now gives us for the velocity components:

$$\int_K \mathbf{v} - \pi_h \mathbf{v} \, dx = 0 \quad \forall K \in \Omega_h,$$
$$\int_e \chi_h \cdot (\mathbf{v} - \pi_h \mathbf{v}) \, do = 0 \quad \forall e \in \Omega_h, \ \forall \chi_h \in P^1(e)$$

We construct the operator π_h as:

$$\pi_h := C_h + E_h + B_h,$$

where C_h denotes the Clement-Interpolation (determines the outer degrees of freedom), E_h is the mean value on the edges (1 degree of freedom for each edge) and B_h defines the mean value in each element. Altogether we have $6 + 1 = 7$ degrees of freedom for the triangular element and 9 for the rectangular element. For each edge we can only set one value, the pressure ξ_h however is linear on the edge, so determines two conditions. If we use instead of C_h the nodal interpolation I_h, it holds:

$$\int_e \chi_h B_h \mathbf{v} \, do = \int_e \chi_h (\mathbf{v} - I_h \mathbf{v}) \, do \quad \forall \chi_h \in P^1.$$

Using the fact that the interpolation error vanishes in the corners $v(x_i) = I_h v(x_i)$, the edge condition is well defined.

The $Q^2 - P^{1,dc}$ and the $P^{2,b} - P^{1,dc}$ element are *inf-sup* stable and order optimal $O(h^2)$.

d) Further conformal elements with discontinuous pressure. In general, the elements $P^k - P^{k-2,dc}$ as well as the elements $Q^k - P^{k-1,dc}$ are *inf-sup* stable for $k \geqslant 2$. The $Q^2 - P^{1,dc}$ element is used very often because of the advantages: local mass conservation due to discontinuous pressure and optimal convergence order for sufficient regularity:

$$\|\nabla(\mathbf{v} - \mathbf{v}_h)\| + \|p - p_h\| = O(h^2)$$

The *mass matrix* in the pressure ansatz space is a block diagonal matrix without coupling between different elements. This will play a role in the discussion of suitable solution methods. Further, the enriched $P^{k,b} - P^{k-1,dc}$ elements are stable. Stability can be shown with Lemma 3.13 using a construction of a suitable projection operator. The generalization to three spatial dimension has to be treated seperately for most elements. Often, the enrichment with several *bulb*-functions is necessary.

3.2.2 Conformal spaces with continuous pressure

If we choose for the pressure space $L_h \subset L_0^2(\Omega)$ a continuous space $L_h \subset C(\Omega)$, we do not have local mass conservation. These spaces have a smaller number of degrees of freedom (and therefore a smaller solution complexity) and have the practical advantage that velocity and pressure spaces can be treated similar. In Figure 3.2 we show some of them.

a) The $P^1 - P^1$ and the $Q^1 - Q^1$ element. These elements do not fulfill the *inf-sup* condition. This holds true for all *equal-order elements* with the same approximation degree in the velocity and pressure. The proof can be done using a non-trivial kernel of the operator grad as in the case of the $Q^1 - P^{0,dc}$. However, the proof is more involved, since there is no more a unique mapping from a pressure test function to one element each.

b) The Mini-element $P^{1,b} - P^1$. Following the argumentation of the $P^{1,b} - P^{0,dc}$ element, we enrich the velocity with one *bulb*-degree of freedom:

$$V_h := [P^1]^d \oplus \operatorname{span}\left\{ \begin{pmatrix} \beta_K^x \\ \beta_K^y \end{pmatrix} b_K, K \in \Omega_h \right\}.$$

This element is called *Mini-element*. The stability proof follows again with the Fortin criterion Lemma 3.13. Due to the continuity of the pressure space, we now have for $p_h \in L_h$ and $\mathbf{v}_h \in V_h$:

$$(p_h, \nabla \cdot \mathbf{v}_h) = -(\nabla q_h, \mathbf{v}_h).$$

For the projection operator π_h the additional condition

$$(\nabla q_h, \mathbf{v} - \pi_h v) = 0$$

has to be fulfilled. For this, the additional two *bulb*-degrees of freedom (β_K^x, β_K^y) are available for each element. The finite element approximation with the Mini-element converges linear. The enriched $P^{k,b} - P^{k-1}$ elements are *inf-sup* stable.

c) The Taylor-Hood elements $P^2 - P^1$ and $Q^2 - Q^1$. These elements are used very often. They are *inf-sup* stable. The proof of the stability is very involved. The Taylor-Hood elements can be defined immediately on tetrahedrons and cubes in three dimension. The generalization $Q^k - Q^{k-1}$ is stable for each $k \geqslant 2$. The triangular elements $P^k - P^{k-2}$ is stable for $k \geqslant 3$ with order difference of 2.

d) The $Q^1 -$ iso Q^1-element. This class of elements takes a special role: the velocity space V_h is the spcae of piecewise bi-linear functions on the mesh Ω_h. Thre pressure space is build from piecewise bi-linear functions on the mesh Ω_{2h}, i.e. on *patches* of elements. This ansatz is *inf-sup* stable and the convergence order

$$\|\nabla(\mathbf{v} - \mathbf{v}_h)\| + \|p - p_h\| \leqslant ch(\|\nabla^2 \mathbf{v}\| + \|\nabla p\|).$$

holds. This is not optimal. Because of the definition of the pressure on macro elements, this ansatz produces many matrix entries, comparable with the $Q^2 - Q^1$ Taylor Hood element, but delivers a lower convergence rate. However, this space will play a role in the construction of *stabilized finite elements*. In general, the spaces $Q^k -$ iso Q^k fulfill the *inf-sup condition*, likewise the triangular elements $P^k -$ iso P^k and the generalization to the three dimensional case. Again, the proof of the inf-sup stability can be derived with local arguments using the Fortin criterion.

3.2.3 Non-conformal elements

In order to enable finite element ansatz spaces with preferably low polynomial degree, non-conformal elements play a role. The lowest order elements $P^1 - P^{0,dc}$ and $Q^1 - P^{0,dc}$, respectively, turn out to be not inf-sup stable. Here, the velocity space is too small in comparison to the pressure space.

Let us illustrate this in the case of the $Q^1 - P^{0,dc}$ element: let us assume that the mesh consists of N rectangles. Each rectangle has on degree of freedom for the pressure (i.e. N total) and 4 degrees of freedom for the velocity. Each velocity degree of freedom is shared with the four adjacent elements. This means, there are $2N$ degrees of freedom in total for the velocity.

Instead of demanding continuity of the velocity $\mathbf{v}_h$ in the corners (and by this along the whole boundaries), we prescribe nodal functionals at the middle points of the edges. The resulting function is continuous in those points but not in general along the whole edge, i.e. $\mathbf{v}_h \notin H_0^1(\Omega)^d$. Each rectangle has 4 degrees of freedom for the velocity, but those are shared only with 2 adjacent elements. Altogether, we obtain $4N$ velocity degrees of freedom and N pressure degrees of freedom. It turns out that this space actually is inf-sup stable.

On rectangles, we define a local ansatz space through the basis given by

$$Q^{1,rot} = \operatorname{span}\{1, x, y, x^2 - y^2\}.$$

The special basis function $x^2 - y^2$ instead of xy is necessary for the construction of an unisolvent ansatz. It holds for all four nodes $x_{ij} = \{(-1)^i, (-1)^j\}$ of the rectangle $K = (-1, 1)^2$ that $x \cdot y = 0$. This element results from a rotation of the Q^1-ansatz by $45°$.

Due to the non-conformity, a lot of difficulties have to be tackled. We have to define differentiation operators piecewise:

$$(\nabla_h \mathbf{v}_h)_{|K} := \nabla(\mathbf{v}_{h|K}), \quad \forall K \in \Omega_h.$$

We look for a solution $\mathbf{v}_h \in V_h$ and $p_h \in L_h$ with

$$\begin{aligned} (\nabla_h \mathbf{v}_h, \nabla_h \phi_h) - (p_h, \nabla_h \cdot \phi_h) &= (f, \phi_h) \quad \forall \phi_h \in V_h \\ (\nabla_h \cdot \mathbf{v}_h, \xi_h) &= 0 \qquad\quad \forall \xi_h \in L_h. \end{aligned}$$

The most prominent non-conformal Stokes elements are the $P^{1,nc} - P^{0,dc}$ and the $Q^{1,rot} - P^{0,dc}$ element. The latter one is called *rotated bilinear* or *Rannacher-Turek element*, see Figure 3.2. For both elements, the degrees of freedom are prescribed on the middle of each edge. Thus, the velocity space is bigger than in the case of the conformal $P^1 - P^{0,dc}$ and $Q^1 - P^{0,dc}$-elements since each edge is shared by only two elements.

Theorem 3.15. The non-conformal $P^{1,nc} - P^{0,dc}$-element as well as the rotated-bilinear $Q^{1,rot} - P^{0,dc}$-element are inf-sup stable.

Proof. A detailed proof is an exercise. According to the Fortin criterion, one has to define an interpolation operator for which the orthogonality property

$$(\xi_h, \nabla \cdot (\mathbf{v} - \pi_h \mathbf{v})) \quad \forall \xi_h \in L_h,$$

is fulfilled. Technical differences will appear since for $\pi_h \mathbf{v} \in V_h$ no conformity holds true. □

Theorem 3.16 (Non-conformal Stokes elements)**.** The non-conformal Stokes elements $P^{1,nc} - P^{0,dc}$ as well as $Q^{1,rot} - P^{0,dc}$ fulfill the *inf-sup* condition with a constant independent of h. If $v \in H^2(\Omega)^d$ and $p \in H^1(\Omega)$ the following error estimates hold true:

$$\|\nabla(\mathbf{v} - \mathbf{v}_h)\| + \|p - p_h\| \leqslant ch(\|\nabla^2 \mathbf{v}\| + \|\nabla p\|),$$

and

$$\|\mathbf{v} - \mathbf{v}_h\| \leqslant ch^2(\|\nabla^2 \mathbf{v}\| + \|\nabla p\|).$$

Proof. We only give a sketch of the proof. The inf-sup condition was already proven in Theorem 3.15.

The proof of the error estimate is difficult due to the lack of the Galerkin orthogonality. Instead of Lemma 3.7 it only holds for the error $\mathbf{e}_v := \mathbf{v} - \mathbf{v}_h$ and $e_h^p := p - p_h$:

$$(\nabla_h \mathbf{e}_h^v, \nabla_h \phi_h) - (e_h^p, \nabla_h \cdot \phi_h) = (\nabla \mathbf{v}, \nabla_h \phi_h) - (p, \nabla_h \cdot \phi_h) - (\mathbf{f}, \phi_h).$$

We define the rest term

$$R^{nc}(\mathbf{v}, p) := \max_{\phi_h \in V_h} \frac{(\nabla \mathbf{v}, \nabla_h \phi_h) - (p, \nabla_h \cdot \nabla \phi_h) - (\mathbf{f}, \phi_h)}{\|\nabla_h \phi_h\|}.$$

This conformity error does not vanish and has to be estimated. Using integration by parts and $-\Delta \mathbf{v} + \nabla p = f$ (in the L^2-sense) it holds:

$$(\nabla \mathbf{v}, \nabla_h \phi_h) - (p, \nabla_h \cdot \phi_h) - (\mathbf{f}, \phi_h) = \sum_{K \in \Omega_h} \int_{\partial K} (\partial_n \mathbf{v} - p\mathbf{n}) \cdot \phi_h, do.$$

Jede Kante $e \in \partial K$ ist entweder gemeinsame Kante von zwei benachbarten Elementen oder Kante auf dem Rand des Gebietes $e \in \partial\Omega$. Diese "au"seren Kanten m"ussen wir nicht weiter beachten, da hier im Fall von Dirichlet-Randwerten $\phi = 0$ gilt.

Es bleiben die inneren Kanten, die stets als Paar auftauchen. Wir gehen davon aus, dass $\nabla \mathbf{v}$ und p entlang der Kanten stetig sind. Dann gilt wegen $\mathbf{n} = -\mathbf{n}'$ von den beiden Seiten:

$$(\nabla \mathbf{v}, \nabla_h \phi_h) - (p, \nabla_h \cdot \phi_h) - (\mathbf{f}, \phi_h) = \sum_{e \in \Omega_h} \int_{\partial K} (\partial_n \mathbf{v} - p\mathbf{n})[\phi_h] \, do,$$

wobei wir mit $[\phi_h]$ den Sprung der (unstetigen) Testfunktion "uber den Rand bezeichnen:

$$[\phi_h](x_e) = \lim_{h \downarrow 0} \phi_h(x_e + \mathbf{n}h) - \phi_h(x_e - \mathbf{n}h), \quad x_e \in e.$$

Die Funktion ϕ_h ist auf e linear und stetig im Mittelpunkt, es gilt also

$$\int_e [\phi_h] \, do = 0.$$

Wir k"onnen somit im Skalarprodukt beliebige konstante Funktionen einf"ugen, etwa $\mathcal{P}_e(\partial_n \mathbf{v} - p\mathbf{n})$, die L^2-Projektion von $\partial_n \mathbf{v} - p\mathbf{n}$ auf den Mittelwert ("uber die Kante e):

$$\int_e (\partial_n \mathbf{v} - p\mathbf{n}) \cdot [\phi_h] \, do = \int_e \big((\partial_n \mathbf{v} - p\mathbf{n}) - \mathcal{P}_e(\partial_n \mathbf{v} - p\mathbf{n})\big) \cdot \big([\phi_h] - \mathcal{P}_e[\phi_h]\big) \, do$$

Mit der Fehlerabsch"atzung f"ur die L^2-Projektion gilt

$$\int_e \big((\partial_n \mathbf{v} - p\mathbf{n}) - \mathcal{P}_e(\partial_n \mathbf{v} - p\mathbf{n})\big) \cdot \big([\phi_h] - \mathcal{P}_e[\phi_h]\big) \, do \leqslant ch\big(\|\nabla^2 v\|_{K_1 \cup K_2} + \|\nabla p\|_{K_1 \cup K_2}\big) \|\nabla \phi_h\|_{K_1 \cup K_2},$$

wobei K_1 und K_2 die an e grenzenden Zellen sind. Zusammen folgt f"ur den Konformit"atsfehler

$$R^{nc}(\mathbf{v}, p) \leqslant ch,$$

also ein zus"atzlicher Fehler, welcher den Energiefehler nicht dominiert.

Auf Basis dieser gest"orten Galerkin-Orthogonalit"at kann die A priori Absch"atzung f"ur Energienorm und L^2-Norm nachgewiesen werden.

□

3.2.4 Praktische Aspekte verschiedener Stokes-Elemente

Allgemein betrachten wir nun Finite-Elemente Paare $V_h \times L_h$, gegeben mit Basisfunktionen

$$V_h := \operatorname{span}\{\phi_h^i, \ i = 1, \dots, N_h^v\}, \quad L_h := \operatorname{span}\{\xi_h^i, \ i = 1, \dots, N_h^p\}.$$

Die Systemmatrix kann wird "ublicherweise in Blockform geschrieben als

$$\mathbf{A}_h := \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ -\mathbf{B}^T & 0 \end{pmatrix}.$$

Dabei sind die einzelnen Matrixbl"ocke $\mathbf{A} \in \mathbb{R}^{N_h^v \times N_h^v}$ sowie $\mathbf{B} \in \mathbb{R}^{N_h^v \times N_h^p}$ d"unn besetzt. Es zeigt sich, dass auf einem gegebenen Gitter die Dimension der Matrizen aber insbesondere die Anzahl der von Null verschiedenen Eintr"age in der Matrix stark variiert. Im Folgenden werden wir die Vernetzungsstruktur der Matrix f"ur einige Finite-Elemente Paare genauer untersuchen. Hierzu betrachten wir exemplarisch zwei einfache Tensorprodukt-Gitter, wie in Abbildung 3.6 dargestellt. Die Gitter bestehen jeweils aus $N \times N$ Knoten und somit $(N-1)^2$ Vierecken oder $2(N-1)^2$ Dreiecken.

Die nicht-stabile $P^1 - P^1$ sowie $Q^1 - Q^1$-Elemente F"ur die Dimensionen der R"aume gilt $N_h^v = 2N$ sowie $N_h^p = N$. Wir beginnen mit der Matrix $\mathbf{A}$. Die Knotenbasisfunktion ϕ_h^i koppelt jeweils mit sich selbst, sowie mit den angrenzenden 8 Knoten im Fall eines Viereckgitters und den 6 Knoten im Fall des Dreiecksgitters. Weiter existieren zu jedem Gitterpunkt jeweils zwei Basisfunktionen f"ur die beiden Komponenten der Geschwindigkeit $(d = 2)$. Wir gehen hier davon aus, dass die beiden Geschwindigkeitskomponenten "uberall

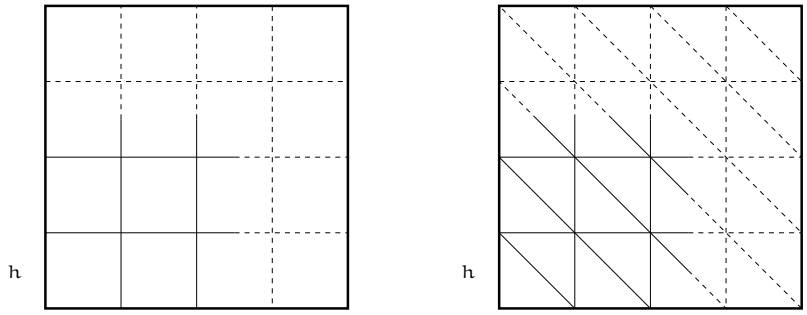


Figure 3.6: Strukturierte Tensorprodukt Gitter bestehend aus N^2 Knoten und $(N-1)^2$ Vierecken, bzw. $2(N-1)^2$ Dreiecken.

miteinander koppeln. Dies ist bei den Stokes-Gleichungen nicht der Fall, jedoch sp"ater bei den nichtlinearen Navier-Stokes Gleichungen. Somit koppelt jeder der $2N$ Geschwindigkeitsfreiheitsgrade in $\mathbf{A}$ mit 18 Freiheitsgraden auf Viereckgittern und 14 Freiheitsgraden in Dreiecksgittern. Hinzu kommen in der Matrix $\mathbf{B}$ weitere 9, bzw. 7 Kopplungen zu den benachbarten Dr"ucken. Insgesamt hat die Matrix $\mathbf{A}_h$ somit $9N^2$ Eintr"age mit insgesamt $72\,N$ von Null verschiedenen Eintr"agen im Falle eines Viereckgitters und $56\,N$ auf Dreiecksgittern.

Das stabile $P^{1,b}-P^1$ Mini-Element Hier vergr"o"sert sich die Dimension des Geschwindigkeitsraums zu $N_h^v = 2N + 4(N-1) = 6N + O(1)$. Jeder der zus"atzlichen $4(N-1)^2$ Freiheitsgrade koppelt zu den 6 Geschwindigkeiten und 3 Dr"ucken im Dreieck. Somit ergibt sich die Gesamtdimension von $\mathbf{A}_h$ mit $6N^2 + O(N)$ Zeilen und Spalten bei insgesamt $52N^2 + 26N^2 + O(N) = 78N^2 + O(N)$ von Null verschiedene Eintr"age.

Das stabile $Q^2-P^{1,dc}$ Element Die Matrix $\mathbf{A}$ hat $2N^2$ Geschwindigkeitsfreiheitsgrade auf Gitterknoten, $2(N-1)^2 = 2N^2 + O(N)$ Freiheitsgrade in Zellmitten und $4N(N-1) = 4N^2 + O(N)$ Freiheitsgrade auf den Kannten. Wir bestrachten stets nur die wesentliche Ordnung, vereinfachen also zu $N_h^v = 8N^2$. Betrachtet man die jeweiligen Tr"ager der Basisfunktionen, so koppeln die $2N^2$ Eckenfreiheitsgrade zu $2 \cdot 25$ weiteren Geschwindigkeitsbasisfunktionen, die $2N^2$ inneren Freiheitsgrade zu $2 \cdot 9$ Basisfunktionen und schlie"slich die $4N^2$ Kanntenfreiheitsgrade zu $2 \cdot 15$. Zusammen ergeben sich f"ur die Matrix $\mathbf{A}$ insgesamt $256N^2$ von Null verschiedene Eintr"age.

Hier ist der Raum L_h unstetig mit 3 Freiheitsgraden pro Viereck. Es gilt somit $N_h^p = 3(N-1)^2 = 3N^2 + O(N)$. Wir z"ahlen die Kopplungen: Jeder der $2N^2$ Eck-Freiheitsgrade des Geschwindigkeitsraums koppelt zu $4 \cdot 3$ Druckfunktionen. Die $2N^2$ inneren zu 3 Basisfunktionen und die $4N^2$ Kanntenfunktionen zu 6 Freiheitsgraden im Druckraum. Die Matrix $\mathbf{B}$ hat somit $54N^2$ von Null verschiedene Elemente.

Insgesamt ergibt sich f"ur $\mathbf{A}_h$ die Dimension $11N^2 \times 11N^2$ bei $364N^2 + O(N)$ von Null verschiedenen Eintr"agen.

Das stabile $Q^{1,rot} - P^0$ Element Abschlie"send behandeln wir noch einen nichtkonformen Fall. Hier verf"ugt der Geschwindigkeitsraum "uber $N_h^v = 4N(N-1) = 4N^2 + O(N)$ Freiheitsgrade auf den Kannten. Jede dieser Basisfunktionen koppelt mit 14 weiteren Geschwindigkeitsfreiheitsgraden und mit 2 der insgesamt $N_h^p = (N-1)^2 = N^2 + O(N)$ Druckfunktionen. Somit ergibt sich eine Matrix $\mathbf{A}_h$ der Dimension $5N^2 \times 5N^2$ mit $72N^2$ von Null verschiedenen Eintr"agen.

3.3 Stabilized finite elements for the Stokes equations

A problem of the finite element pairs for the Stokes equations which we have considered so far is the relatively high effort for the construction and administration of stable ansatz spaces. The inf-sup condition has to hold true and the velocity and pressure degrees of freedom have to be handled differently. However, the choice of ansatz spaces of the same order for the velocity and the pressure leads to a non inf-sup stable discretization. Such an *equal-order*-discretization, however, would simplify the implementation a lot since all solution components v_h^i for $i = 1, \dots, d$ and p_h could be represented with the same finite element basis ϕ_h^i for $i = 1, \dots, N$ as

$$\mathbf{u}_h = \{\mathbf{v}_h, p_h\}, \quad \mathbf{u}_h = \sum_{i=1}^{N} \mathbf{u}_i \phi_h^i, \quad \mathbf{u}_i \in \mathbb{R}^{d+1}.$$

In the Fortin criterion Lemma 3.13, the operator $\pi_h : \mathcal{V} \to V_h$ needs to fulfill H^1 stability and an orthogonality property. This is not possible for equal-order elements. We consider the critical part in the proof and formulate a Clement interpolation for $p_h \in L_h \subset \mathcal{L}$ with conformal and globally continuous ansatz spaces

$$\begin{aligned} \gamma \|p_h\| &\leqslant \sup_{\phi \in \mathcal{V}} \frac{(p_h, \nabla \cdot (\phi - C_h\phi))}{\|\nabla \phi\|} + \sup_{\phi \in \mathcal{V}} \frac{(p_h, \nabla \cdot C_h\phi)}{\|\nabla C_h \phi\|} \frac{\|\nabla C_h \phi\|}{\|\nabla \phi\|} \\ &\leqslant \sup_{\phi \in \mathcal{V}} \sum_{K \in \Omega_h} \frac{(\nabla p_h, \phi - C_h\phi)_K}{\|\nabla \phi\|} + c \sup_{\phi_h \in V_h} \frac{(p_h, \nabla \cdot \phi_h)}{\|\nabla \phi_h\|}. \end{aligned} \tag{3.15}$$

The boundary terms of the integration by parts vanish due to the continuity of V_h and L_h. Using the interpolation property for the Clement interpolation it follows

$$\begin{aligned} \gamma \|p_h\| &\leqslant c \sup_{\phi_h \in V_h} \frac{(p_h, \nabla \cdot \phi_h)}{\|\nabla \phi_h\|} + c \sup_{\phi \in V} \sum_{K \in \Omega_h} h_K \frac{\|\nabla p_h\| \, \|\nabla \phi\|_{P_K}}{\|\nabla \phi\|} \\ &\leqslant c \sup_{\phi_h \in V_h} \frac{(p_h, \nabla \cdot \phi_h)}{\|\nabla \phi_h\|} + \tilde{c} \left(\sum_{K \in \Omega_h} h_K^2 \|\nabla p_h\|_K^2 \right)^{\frac{1}{2}}, \end{aligned}$$

where h_K denotes the diameter of the element $K \in \Omega_h$.

Lemma 3.17 (Modified inf-sup condition)**.** Let $V_h \times L_h \subset \mathcal{V} \times \mathcal{L}$ be a finite element pair with continuous pressure. Then, for $p_h \in L_h$ the modified inf-sup condition

$$\gamma_h \|p_h\| \leqslant \sup_{\phi_h \in V_h} \frac{(p_h, \nabla \cdot \phi_h)}{\|\nabla \phi_h\|} + \left(\sum_{K \in \Omega_h} h_K^2 \|\nabla p_h\|_K^2 \right)^{\frac{1}{2}}, \tag{3.16}$$

holds true with a constant $\gamma_h = \gamma/(\tilde{c})$.

This Lemma 3.17 is the basis for *stabilized finite element methods*. For the proof of existence of a unique solution we consider the approach of using a problem specific norm $|||\cdot|||$ (compare Lemma 2.30) and choose here:

$$|||u_h||| := \left(\|\nabla \mathbf{v}_h\|^2 + \|p_h\|^2 + \sum_{K \in \Omega_h} h_K^2 \|\nabla p_h\|_K^2 \right)^{\frac{1}{2}}.$$

Before we can introduce stabilized finite elements, we require a further auxiliary result:

Lemma 3.18 (Inverse estimation)**.** For $p_h \in L_h$ with $\int_\Omega p_h \, dx = 0$ the following estimation holds true

$$\sum_{K \in \Omega_h} h_K^2 \|\nabla p_h\|_K^2 \leqslant c \|p_h\|^2.$$

Proof. The proof uses the equivalence of norms in finite dimensional spaces. Using the transformation to reference elements $T_K : \hat{K} \to K$ we get:

$$\sum_{K \in \Omega_h} h_K^2 \int_K |\nabla p_h(x)|^2 \, dx = \sum_{K \in \Omega_h} h_K^2 \int_{\hat{K}} \det(\hat{\nabla} T_K) |\nabla p_h(x)|^2 \, d\hat{x}.$$

With $x = T_K(\hat{x})$ and $\hat{p}_h(\hat{x}) = p_h(x)$ it holds

$$\sum_{K \in \Omega_h} h_K^2 \int_K |\nabla p_h(x)|^2 \, dx = \sum_{K \in \Omega_h} h_K^2 \int_{\hat{K}} \det(\hat{\nabla} T_K) |\hat{\nabla} T_K^{-T} \hat{\nabla} \hat{p}_h|^2 \, d\hat{x}.$$

For regular finite element grids it holds

$$\det(\hat{\nabla} T_K) \approx h_K^d, \quad |\hat{\nabla} \hat{T}_K^{-T}| \sim |\hat{\nabla} \hat{T}_K|^{-1} \approx h_K^{-1}.$$

Together, this leads to

$$\sum_{K \in \Omega_h} h_K^2 \int_K |\nabla p_h(x)|^2 \, dx \leqslant c \sum_{K \in \Omega_h} h_K^d \|\hat{\nabla} \hat{p}_h\|_{\hat{K}}^2.$$

Further by expanding the seminorm to a norm and using the norm equivalence on finite dimensional spaces, we have

$$\sum_{K\in\Omega_h} h_K^d \|\hat{\nabla}\hat{p}_h\|_{\hat{K}}^2 \leqslant \sum_{K\in\Omega_h} h_K^d \|\hat{p}_h\|_{H^1(\hat{K})}^2 \leqslant c \sum_{K\in\Omega_h} h_K^d \|\hat{p}_h\|_{L^2(\hat{K})}^2.$$

The result follows by back transformation on K with $\det(\hat{\nabla}T_K^{-1}) \approx h_K^{-d}$. □

Theorem 3.19 (Stability of the modified Stokes equations)**.** Let $V_h \times L_h \subset \mathcal{V} \times \mathcal{L}$ be a conformal finite element pair. Then, the modified Stokes problem

$$\begin{aligned} (\nabla \mathbf{v}_h, \nabla \phi_h) - (p_h, \nabla \cdot \phi_h) &= (f, \phi_h) \quad \forall \phi_h \in V_h, \\ (\nabla \cdot \mathbf{v}_h, \xi_h) + s_h(\xi_h, p_h) &= 0 \qquad\quad \forall \xi_h \in L_h. \end{aligned} \tag{3.17}$$

with stabilizing term $s_h(\xi_h, p_h) := \alpha \sum_{K\in\Omega_h} h_K^2 (\nabla p_h, \nabla \xi_h)_K$ has for each $\alpha > 0$ a unique solution $(\mathbf{v}_h, p_h) \in V_h \times L_h$. The stability condition holds:

$$\|\nabla \mathbf{v}_h\| + \|p_h\| + \left(\alpha \sum_{K\in\Omega_h} h_K^2 \|\nabla p_h\|^2 \right)^{\frac{1}{2}} \leqslant c\|f\|.$$

Proof. Existence of a solution follows from the injectivity. Testing with $\phi_h := \mathbf{v}_h$ and $\xi_h := p_h$ and adding both equations leads to the estimation

$$\begin{aligned} \|\nabla \mathbf{v}_h\|^2 + \alpha \sum_{K\in\Omega_h} h_K^2 \|\nabla p_h\|_K^2 = (f, \mathbf{v}_h) \leqslant c_p \|\mathbf{f}\| \, \|\nabla \mathbf{v}_h\| \leqslant \frac{c_p^2}{2}\|\mathbf{f}\|^2 + \frac{1}{2}\|\nabla \mathbf{v}_h\|^2 \\ \Leftrightarrow \quad \|\nabla \mathbf{v}_h\|^2 + 2\alpha \sum_{K\in\Omega_h} h_K^2 \|\nabla p_h\|_K^2 \leqslant c_p^2 \|\mathbf{f}\|^2. \end{aligned}$$

Using the auxiliary result Lemma 3.18 it holds

$$|||U_h|||^2 = \|\nabla \mathbf{v}_h\|^2 + \|p_h\|^2 + \sum_{K\in\Omega_h} \alpha h_K^2 \|\nabla p_h\|_K^2 \leqslant c\|\mathbf{f}\|^2.$$

For $\mathbf{f} = 0$ it follows $U_h = \{\mathbf{v}_h, p_h\} = 0$, i.e. uniqueness of a solution. □

Using piecewise linear finite elements for the velocity and the pressure, we obtain optimal order a priori estimates:

Lemma 3.20 (A priori error of the modified Stokes equations)**.** For the modified Stokes problem (3.17) with $p \in H^1(\Omega)$ and $\mathbf{v} \in H^2(\Omega)$ the following a priori estimates hold true

$$\begin{aligned} \|\nabla(\mathbf{v} - \mathbf{v}_h)\| + \|p - p_h\| &\leqslant ch(\|\nabla^2 \mathbf{v}\| + \|\nabla p\|) \\ \|\mathbf{v} - \mathbf{v}_h\| &\leqslant ch^2(\|\nabla^2 \mathbf{v}\| + \|\nabla p\|). \end{aligned}$$

Proof. The proof is a modification of Lemma 3.7. Note that we here have a modified Galerkin orthogonality in the divergence equation

$$(\nabla \cdot (\mathbf{v} - \mathbf{v}_h), \xi_h) = \alpha \sum_{K \in \Omega_h} h_K^2 (\nabla p_h, \nabla \xi_h)_K.$$

□

Due to the stabilization term there appears an error term of first order which does not depend on the degree of the finite element ansatz spaces V_h, L_h, but comes from the modification of the Stokes equation. For ansatz spaces of higher order, e.g. $P^2 - P^2$ the modified Stokes equation has a unique stable solution, but the convergence is limited by the linear order term.

Another weakness of the stabilized form is the introduction of a higher regularity for the pressure. The stabilization term also leads to an unphysical boundary condition:

$$\sum_{K \in \Omega_h} \alpha h_K^2 (\nabla p_h, \nabla \xi_h)_K = \sum_{K \in \Omega_h} \alpha h_K^2 \Big\{ -(\Delta p_h, \xi_h)_K + \int_{\partial K} \partial_n p_h \xi_h \, do \Big\}.$$

The boundary terms vanish within the domain, but do not vanish on the boundary of the domain $\partial\Omega$. This artificial boundary condition falsifies the flow profile, see Figure 3.7. For $h_K \to 0$ the influence on the flow reduces.

Further, the correct choice of the stabilization parameter α is often critical. The parameter influences the stability of the method - and has therefore influence on the convergence of linear solvers. It also determines the constant in the error estimation. The correct dependency of α on the viscosity can be computed as follows. Let us assume that α_0 is a suitable parameter for the case $\nu = 1$:

$$\begin{aligned} (\nabla \mathbf{v}_h, \nabla \phi_h) - (p_h, \nabla \cdot \phi_h) &= (\mathbf{f}, \phi_h) \\ (\nabla \cdot \mathbf{v}_h, \xi_h) + \alpha_0 \sum_{K \in \Omega_h} h_K^2 (\nabla p_h, \nabla \xi_h) &= 0. \end{aligned}$$

Multiplication of the first equation with ν gives with $\bar{p}_h := \nu p_h$ and $\bar{\mathbf{f}} := \nu \mathbf{f}$:

$$\begin{aligned} \nu(\nabla \mathbf{v}_h, \nabla \phi_h) - (\bar{p}_h, \nabla \cdot \phi_h) &= (\bar{\mathbf{f}}, \phi_h) \\ (\nabla \cdot \mathbf{v}_h, \xi_h) + \frac{\alpha_0}{\nu} \sum_{K \in \Omega_h} h_K^2 (\nabla \bar{p}_h, \nabla \xi_h) &= 0. \end{aligned}$$

The correct scaling of the stabilization is therefore $\frac{h_K^2}{\nu}$.

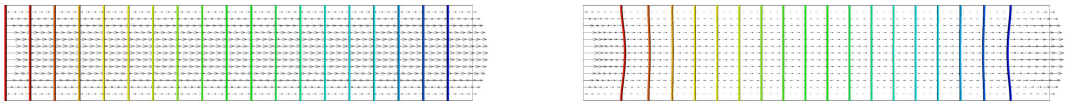


Figure 3.7: Channel flow with stable finite elements (left) and with the stabilized form (right). Along with the velocity vectors, the pressure isolines are illustrated.

3.3.1 The consistent *PSPG*-form

The consistency error in the pressure stabilization is of first order and will dominate the error if higher order finite elements are used.

We can tackle this issue by considering a *fully consistent* stabilization approach. We consider the following stabilized Stokes problem:

$$\begin{aligned} (\nabla \mathbf{v}_h, \nabla \phi_h) - (p_h, \nabla \cdot \phi_h) &= (f, \phi_h) \quad \forall \phi_h \in V_h, \\ (\nabla \cdot \mathbf{v}_h, \xi_h) + s_h(\xi_h, p_h) &= g_h(\xi_h) \quad \forall \xi_h \in L_h, \end{aligned} \tag{3.18}$$

with stabilizing term $s_h(\xi_h, p_h) := \alpha \sum_{K \in \Omega_h} h_K^2 (\nabla p_h, \nabla \xi_h)_K$ and the form

$$g_h(\xi_h) := \alpha \sum_{K \in \Omega_h} h_K^2 (f + \Delta \mathbf{v}_h, \nabla \xi_h)_K.$$

If we insert the continuous solution $\{\mathbf{v}, p\} \in C^2(\Omega)^d \times C^1(\Omega)$ in the second equation of (3.18), the stabilization terms vanish. It can be shown that this approach leads to a stable discretization of the Stokes problem for all equal order ansatz spaces.

This stabilization method can also be derived in form of a Petrov-Galerkin method by testing the classical Stokes problem with the test function

$$\tilde{\phi}_h := \phi_h + \alpha h^2 \nabla \xi_h.$$

The approach is therefore often called *Pressure Stabilized Petrov-Galerkin*, or PSPG. For more details we refer to e.g. [23].

3.3.2 Stabilisierung mit lokalen Projektionen

Eine alternative Stabilisierungsmethode ist die Methode der *Lokalen Projektionen* (LPS) von Becker und Braack [3]. Sie ist eng verbunden mit der PSPG Formulierung. Hier war der konsistente Stabilisierungsterm definiert als:

$$S(U_h, \Phi_h) = \sum_{K \in \Omega_h} \alpha h_K^2 (\nabla p_h, \nabla \xi_h)_K - \sum_{K \in \Omega_h} \alpha h_K^2 (\mathbf{f} + \Delta \mathbf{v}_h, \nabla \xi_h)_K.$$

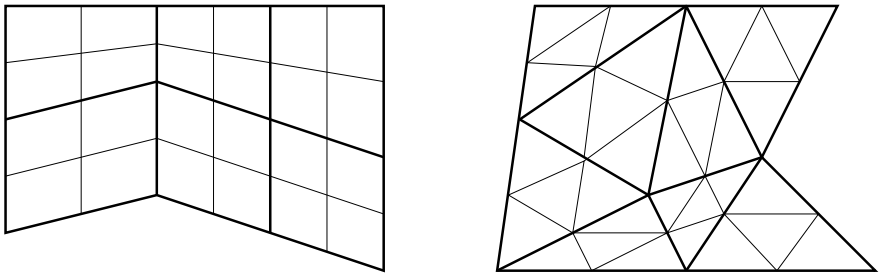


Figure 3.8: Finite Elemente Gitter Ω_h (d"unne Linien) mit zugeh"origem Patchgitter Ω_H (dicke Linien).

Die Idee der Projektionsmethode ist, dass strenge Konsistenz nicht notwendig ist, um eine optimale Fehlerordnung zu erreichen. Stattdessen muss der Konsistenzfehler lediglich klein genug sein, so dass er den gesamten Diskretisierungsfehler nicht dominiert. Weiter beobachten wir, dass Stabilit"at durch den ersten Term, also durch $\sum h_K^2(\nabla p_h, \nabla \xi_h)$ erzeugt wird. F"ur kleine h gilt $\nabla p_h \approx \mathbf{f} + \Delta \mathbf{v}_h$. Statt diesem Konsistenzterm f"ugen wir also eine Approximation an ∇p_h ein, welche wir hier $\overline{\nabla p_h}$ nennen:

$$S(U_h, \Phi_h) = \sum_{K \in \Omega_h} \alpha h_K^2 \Big\{ (\nabla p_h, \nabla \xi_h)_K - (\overline{\nabla p_h}, \nabla \xi_h)_K \Big\}.$$

Wichtig wird lediglich sein, dass der Fehler $\nabla p_h - \overline{\nabla p_h}$ mit richtiger Ordnung in h gegen Null geht. Wir betrachten zun"achst den Fall linearer Elemente in Geschwindigkeit und Druck und w"ahlen f"ur $\overline{\nabla p_h}$ die Projektion von ∇p_h auf die patchweise konstanten Funktionen. Hierzu sei das Gitter in folgender Weise konstruiert:

Definition 3.21 (Gitter mit Patchen)**.** Es sei Ω_h ein Finite Elemente Gitter mit Patch-Struktur: Zu Ω_h existiert ein Grobgitter Ω_H, so dass je m Elemente $K_1, \ldots, K_m \in \Omega_h$ durch regelm"a"sige Verfeinerung eines Elements $P \in \Omega_H$ erzeugt werden.

Abbildung 3.8 zeigt Beispiele f"ur "ubliche Patchgitter.

Wir definieren den Raum der patchweise konstanten Funktionen

$$Q_P := \{\xi \in L^2(\Omega)^d,\ \xi\Big|_P \in \mathbb{R}^d\},$$

und die L^2-Projektion nach Q_P:

$$\overline{\nabla p} \in Q_P : \quad (\overline{\nabla p}, \psi) = (\nabla p, \nabla \psi) \quad \forall \psi \in Q_P.$$

Aufgrund der Orthogonalit"atseigenschaft der L^2-Projektion gilt

$$S(U_h, \Phi_h) := \sum_{P \in \Omega_H} \alpha h_K^2 (\nabla p_h - \overline{\nabla p_h}, \nabla \xi_h)_P = \sum_{P \in \Omega_H} \alpha h_K^2 (\nabla p_h - \overline{\nabla p_h}, \nabla \xi_h - \overline{\nabla \xi_h})_P.$$

F"ur diese erste Form der LPS-Stabilisierung gilt der folgende Satz

Satz 3.22 (LPS-Stabilisierung f"ur lineare Finite Elemente)**.** Es sei $V_h \times L_h$ der Raum der linearen Finiten Elemente. Dann existiert zu jedem $\mathbf{f} \in L^2(\Omega)^d$ eine L"osung von

$$\begin{aligned}(\nabla \mathbf{v}_h, \nabla \phi_h) - (p_h, \nabla \cdot \phi_h) &= (\mathbf{f}, \phi_h) \quad \forall \phi_h \in V_h,\\ (\nabla \cdot \mathbf{v}_h, \nabla \phi_h) + \sum_{P \in \Omega_h} \alpha h_K^2 (\nabla p_h - \overline{\nabla p_h}, \nabla \xi_h - \overline{\nabla \xi_h})_P &= 0 \qquad \forall \xi_h \in L_h.\end{aligned}$$

Diese L"osung erf"ullt die Stabilit"atsabsch"atzung

$$\|\nabla \mathbf{v}_h\|^2 + \|p_h\|^2 + \sum_{K \in \Omega_h} h_K^2 \|\nabla p_h - \overline{\nabla p_h}\|_K \leqslant c\|\mathbf{f}\|^2,$$

sowie auf konvexen oder glatten Gebieten die a priori Absch"atzung

$$\begin{aligned}\|\nabla(\mathbf{v} - \mathbf{v}_h)\| + \|p - p_h\| &\leqslant ch(\|\nabla^2 \mathbf{v}\| + \|\nabla p\|),\\ \|\mathbf{v} - \mathbf{v}_h\| &\leqslant ch^2(\|\nabla^2 \mathbf{v}\| + \|\nabla p\|).\end{aligned}$$

Beweis: Der Beweis zu dieser Aussage findet sich in [3]. Wir verzichten hier auf eine Wiedergabe und werden sp"ater eine abstrakte Variante der LPS-Methode vorstellen. □

Die bisher vorgestellte LPS-Methode ist nicht konsistent, da im allgemeinen $S_h(\cdot, p)(\cdot, p) \neq 0$, sie verf"ugt jedoch "uber st"arkere Konsistenzeigenschaften als die einfache Druckstabilisierung. Im Fall $p \in Q_{2h}$, falls der (kontinuierliche) Druck im Raum der linearen Finiten Elemente auf dem Patchgitter liegt, so verschwindet der Stabilisierungsterm. Dies zeigt sich auch in der zus"atzlichen Norm, bzgl. der die L"osung Stabilit"at aufweist:

$$\sum_K \alpha h_K^2 \|\nabla p - \overline{\nabla p}\|_P^2,$$

welche etwas schw"acher ist, als die Norm der Druckstabilisierung. Schlie"slich f"uhrt die LPS-Methode zu einer geringeren Verf"alschung der L"osung an R"andern. Die Darstellung auf der linken Seite in Abbildung 3.7 korrespondiert zu einer LPS-stabilisierten L"osung.

Im Anschluss betrachten wir nun einen allgemeineren Zugang zur LPS-Methode, welcher es uns auch erm"oglichen wird, Satz 3.22 zu beweisen und sich auf Ans"atze beliebiger Ordnung "ubertragen werden kann. Konsistenz wird wieder nur im schw"acheren Sinne gefordert: Der Konsistenzfehler darf den Approximationsfehler nicht dominieren.

Die allgemeine LPS-Methode basiert auf der Projektion des Drucks $\pi_h : L_h \to \tilde{L}_h$ in einen inf-sup stabilen Ansatzraum $V_h \times \tilde{L}_h$. Der Stabilisierungsterm wird so definiert, dass er den Unterschied zwischen Druck $p_h \in L_h$ und projiziertem, stabilen Druck $\pi_h p_h \in \tilde{L}_h$ "bestraft".

Satz 3.23 (Allgemeiner Stabilit"atssatz f"ur die LPS-Methode)**.** Es sei durch $V_h \times \tilde{L}_h$ ein konformes, inf-sup stabiles Finite Elemente Paar mit inf-sup Konstante $\tilde{\gamma}_h \geqslant \tilde{\gamma} > 0$ gegeben. Weiter existiere ein diskreter, L^2-stabiler Projektionsoperator $\pi_h : L_h \to \tilde{L}_h$

$$\|\pi_h p_h\| \leqslant c_\pi \|p_h\| \quad \forall p_h \in L_h. \tag{3.19}$$

Schlie"slich sei durch $S_h : L_h \times L_h$ eine Stabilisierungsform gegeben, welche den Projektionsfehler beschr"ankt:

$$\|p_h - \pi_h p_h\| \leqslant c_\pi S(p_h, p_h)^{\frac{1}{2}}. \tag{3.20}$$

Dann gilt die modifizierte inf-sup Bedingung

$$\tilde{\gamma}_h \|p_h\| \leqslant \sup_{\phi_h \in V_h} \frac{(p_h, \nabla \cdot \phi_h)}{\|\nabla \phi_h\|} + (1 + \tilde{\gamma}_h c_\pi) S(p_h, p_h)^{\frac{1}{2}},$$

sowie die Stabilit"atsabsch"atzung

$$\sup_{|||\Phi_h|||_h = 1} \left\{ A(U_h, \Phi_h) + S_h(U_h, \Phi_h) \right\} \geqslant c |||U_h|||_h \quad \forall U_h \in V_h \times L_h,$$

mit der Norm

$$|||U_h|||_h^2 := \nu \|\nabla \mathbf{v}_h\|^2 + \gamma_h^2 \|p_h\|^2 + S(p_h, p_h).$$

Beweis: (i) Wir weisen zun"achst die modifizierte inf-sup Bedingung nach. Es sei $p_h \in L_h$ beliebig. Dann gilt mit (3.20)

$$\begin{aligned}
\|p_h\| &\leqslant \|\pi_h p_h\| + \|p_h - \pi_h p_h\| \leqslant \tilde{\gamma}_h^{-1} \sup_{\phi_h \in V_h} \frac{(\pi_h p_h, \nabla \cdot \phi_h)}{\|\nabla \phi_h\|} + c_\pi S(p_h, p_h)^{\frac{1}{2}} \\
&\leqslant \tilde{\gamma}_h^{-1} \sup_{\phi_h \in V_h} \frac{(p_h, \nabla \cdot \phi_h)}{\|\nabla \phi_h\|} + \tilde{\gamma}_h^{-1} \sup_{\phi_h \in V_h} \frac{(p_h - \pi_h p_h, \nabla \cdot \phi_h)}{\|\nabla \phi_h\|} + c_\pi S(p_h, p_h)^{\frac{1}{2}} \\
&\leqslant \tilde{\gamma}_h^{-1} \sup_{\phi_h \in V_h} \frac{(p_h, \nabla \cdot \phi_h)}{\|\nabla \phi_h\|} + (\tilde{\gamma}_h^{-1} + c_\pi) S(p_h, p_h)^{\frac{1}{2}}.
\end{aligned}$$

(ii) Der Nachweis der Stabilit"atsabsch"atzung erfolgt analog zu Satz **??**. Wir testen zun"achst diagonal mit $\Phi_h^1 = \{\mathbf{v}_h, p_h\}$ und erhalten

$$A(U_h, \Phi_h^1) = \nu \|\nabla \mathbf{v}_h\|^2 + S(p_h, p_h).$$

Jetzt w"ahlen wir zu $p_h \in L_h$ ein $\tilde{\mathbf{v}}_h \in V_h$ mit $\|\tilde{\mathbf{v}}_h\| = 1$ und der Eigenschaft

$$\tilde{\gamma}_h \|p_h\| \leqslant (p_h, \nabla \cdot \tilde{\mathbf{v}}_h) + (1 + \tilde{\gamma}_h c_\pi) S(p_h, p_h)^{\frac{1}{2}}.$$

Dann gilt f"ur $\Phi_h^2 := \{-\|p_h\| \tilde{\mathbf{v}}_h, 0\}$ bei mehrfacher Anwendung der Young'schen Ungleichung

$$\begin{aligned}
A(U_h, \Phi_h^2) &= \|p_h\| \Big(-\nu(\nabla \mathbf{v}_h, \nabla \tilde{\mathbf{v}}_h) + (p_h, \nabla \cdot \tilde{\mathbf{v}}_h) \Big) \\
&\geqslant \|p_h\| \Big(\nu \|\nabla \mathbf{v}_h\| + \tilde{\gamma}_h \|p_h\| - (1 + \tilde{\gamma}_h c_\pi) S(p_h, p_h)^{\frac{1}{2}} \Big) \\
&\geqslant \frac{\tilde{\gamma}_h}{4} \|p_h\|^2 - \frac{\nu^2}{\tilde{\gamma}_h} \|\nabla \mathbf{v}_h\|^2 - \frac{(1 + \tilde{\gamma}_h c_\pi)^2}{\tilde{\gamma}_h} S(p_h, p_h).
\end{aligned}$$

Wir kombinieren nun $\Phi_h := \Phi_h^1 + \epsilon \Phi_h^2$, wobei

$$\epsilon := \tilde{\gamma}_h \min \left\{ \frac{1}{2\nu}, \frac{1}{2(1 + \tilde{\gamma}_h c_\pi)^2} \right\}.$$

Dann gilt

$$A(U_h, \Phi_h) + S(U_h, \Phi_h) \geqslant \frac{\nu}{2}\|\nabla \mathbf{v}_h\|^2 + c\tilde{\gamma}_h^2\|p_h\| + \frac{1}{2}S(p_h, p_h)^2,$$

mit $c = \epsilon/\tilde{\gamma}_h$. Im Allgemeinen k"onnen ν und $\tilde{\gamma}_h$ als klein angenommen werden, so dass gilt $c \approx \frac{1}{2}$. Die Ungleichung folgt mit Teilen durch $\|\!|\Phi_h|\!\|_h$. □

Aus diesem Satz folgt als direktes Resultat:

Satz 3.24 (Allgemeiner Existenzsatz f"ur die LPS-Methode)**.** Unter den Voraussetzungen von Satz 3.23 existiert eine eindeutig bestimmte L"osung $\{\mathbf{v}_h, p_h\} \in V_h \times L_h$, welche die folgende Stabilit"atsabsch"atzung erf"ullt:

$$\|\!|U_h|\!\|_h^2 = \nu\|\nabla \mathbf{v}_h\|^2 + \gamma_h^2\|p_h\|^2 + S(p_h, p_h) \leqslant c\nu^{-1}\|\mathbf{f}\|^2.$$

Beweis: Die Existenz einer L"osung folgt aus der Eindeutigkeit, welche unmittelbares Resultat von Satz 3.23 ist. Weiter gilt f"ur diese L"osung

$$c\|\!|U_h|\!\|_h \leqslant \sup_{\|\!|\Phi_h|\!\|_h} \Big\{A(U_h, \Phi_h)+S(U_h, \Phi_h)\Big\} = \sup_{\|\!|\Phi_h|\!\|_h=1} F(\Phi_h) \leqslant c_p\nu^{-\frac{1}{2}}\|\mathbf{f}\| \sup_{\|\!|\Phi_h|\!\|_h=1} \|\!|\Phi_h|\!\|_h.$$

□

Schlie"slich k"onnen wir f"ur die allgemeine LPS-Methode eine abstrakte a priori Absch"atzung f"ur den Diskretisierungsfehler herleiten. Aus technischen Gr"unden ben"otigen wir f"ur diesen Satz eine Cauchy-Schwarz "ahnliche Ungleichung f"ur den Stabilisierungsterm $S(\cdot,\cdot)$. Sp"ater, bei der Analyse konkreter Methoden wird sich diese Ungleichung einfach nachweisen lassen.

Satz 3.25 (Approximation der LPS-Methode)**.** Es gelten die Voraussetzungen von Satze 3.23. Zus"atzlich erlaube der Stabilisierungsterm die Absch"atzung

$$S(p_h, q_h) \leqslant S(p_h, p_h)^{\frac{1}{2}} S(q_h, q_h)^{\frac{1}{2}}. \tag{3.21}$$

Dann gilt f"ur die L"osung der LPS-stabilisierten Stokes-Gleichungen die Approximationseigenschaft

$$\nu^{\frac{1}{2}}\|\nabla(\mathbf{v} - \mathbf{v}_h)\| + \tilde{\gamma}_h\|p - p_h\| \leqslant c\Big(\|\!|U - i_h U|\!\|_h + S_h(i_h p, i_h p)^{\frac{1}{2}}\Big),$$

mit einem Interpolationsoperator $i_h : V \times Q \to V_h \times L_h$.

Beweis: *(i)* Wir teilen den Fehler $E_h := U - U_h$ wieder auf in Interpolationsfehler $U - i_h U$ und Projektionsfehler $i_h U - U_h$. F"ur den letzteren gilt zun"achst mit der Stabilit"atsabsch"atzung

$$\|\!|i_h U - U_h|\!\| \leqslant c \sup_{\|\!|\Phi_h|\!\|_h=1} (A + S)(i_h U - U_h, \Phi_h).$$

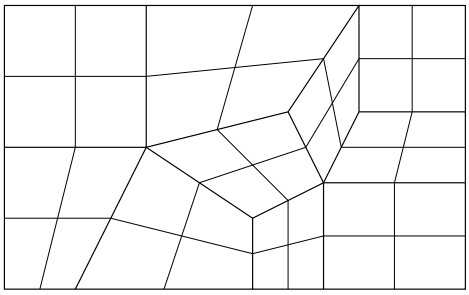


Figure 3.9: Gitter mit Patch-Struktur.

Es gilt nun mit der gest"orten Galerkin-Orthogonalit"at

$$\begin{aligned}(A+S)(i_h u - u_h, \Phi_h) &= (A+S)(u - u_h, \Phi_h) - (A+S)(u - i_h u, \Phi_h)\\ &= S(u, \Phi_h) - S(u - i_h u, \Phi_h) - A(u - i_h u, \Phi_h)\\ &= S(i_h u, \Phi_h) - A(u - i_h u, \Phi_h)\end{aligned}$$

Der zweite Anteil kann unmittelbar abgesch"atzt werden zu

$$A(u - i_h u, \Phi_h) \leqslant |||u - i_h u|||_h \, |||\Phi_h|||_h.$$

F"ur den Stabilisierungsterm gilt bei $\Phi_h = \{\phi_h, \xi_h\}$ mit (3.21)

$$S(i_h u, \Phi_h) = s(i_h p, \xi_h) \leqslant s(i_h p, i_h p)^{\frac{1}{2}} \underbrace{s(\xi_h, \xi_h)^{\frac{1}{2}}}_{\leqslant |||\Phi_h|||_h}.$$

Zusammen ergibt sich die gew"unschte Absch"atzung. □

Im Folgenden geben wir nun konkrete Varianten der LPS-Methode an und "uberpr"ufen die verschiedenen Bedingungen an den Projektionsoperator $\pi_h : L_h \to \tilde{L}_h$ sowie an die Stabilisierungsform.

a) Projektion auf grobes Gitter Es sei ein Gitter Ω_h mit Patch-Struktur Ω_P gegeben, siehe z.B. Abbildung 3.9 oder 3.8. Dabei sei jedes Viereck (f"ur Dreiecke, Tetraeder oder Hexaeder ist eine entsprechende Konstruktion m"oglich) Teil von vier Vierecken, die aus der gemeinsamen regelm"a"sigen Verfeinerung eines Patches entstanden sind. Als Ansatzraum $V_h \times L_h$ auf Ω_h w"ahlen wir den equal-order Raum $Q^r - Q^r$. Als Stabilen Raum betrachten wir das $Q^r - \text{iso}\, Q^r$ Element, also den Raum $V_h \times \tilde{L}_h = V_h \times Q_{2h}$ von gleicher Ordnung. Der Druck ist lediglich auf den gr"oberen Patchen definiert. Hier gilt $Q_{2h} \subset L_h$, der Projektionsoperator ist gerade die Interpolation $\pi_h := i_{2h}$. Dieser ist auf L_h nat"urlich stetig, d.h. Bedingung (3.19) gilt.

Wir w"ahlen den Stabilisierungsterm als

$$S_h(p_h, \xi_h) = \sum_{P \in \Omega_P} \alpha_K h_K^2 (\nabla(p_h - \pi_h p_h), \nabla(\xi_h - \pi_h \xi_h))_P,$$

und m"ussen noch Bedingungen (3.20) sowie (3.21) nachweisen. (3.21) folgt aus der zugrundeliegenden Skalarprodukteigenschaft des Stabilisierungsterms gilt. Mit der inversen Ungleichung gilt:

$$S_h(p_h, p_h) = \sum_{P\in\Omega_P} \alpha_K h_K^2 \|\nabla(p_h - \pi_h p_h)\|_P^2 \geqslant c_{inv}\alpha_0 \sum_{P\in\Omega_P} \|p_h - \pi_h p_h\|_P^2 = c_{inv}\alpha_0 \|p_h - \pi_h p_h\|^2,$$

mit $\alpha_0 = \min\{\alpha_K\}$. Die Inverse Ungleichung darf hier angewendet werden, da aus $\nabla(p_h - \pi_h p_h) = 0$ folgt, dass $p_h - \pi_h p_h$ eine konstante Funktion ist. Dann muss wegen $\pi_h = i_{2h}$ auch gelten, dass $p_h = \pi_h p_h$, also notwendigerweise $p_h - \pi_h p_h = 0$.

Schlie"slich gilt es, den zus"atzlichen Stabilisierungsfehler in der Fehlerabsch"atzung aus Satz 3.25 zu analysieren. Es ist

$$S_h(i_h p, i_h p) = \sum_{P\in\Omega_h} \alpha_K h_K^2 \|\nabla(i_h p - \pi_h i_h p)\|_P^2.$$

Bei hinreichender Regularit"at $p \in H^2(\Omega)$ kommutieren die beiden Interpolationsoperatoren i_h und $\pi_h = i_{2h}$, so dass mit der Stabilit"atsabsch"atzung der Interpolation folgt

$$S_h(i_h p, i_h p) = \sum_{P\in\Omega_h} \alpha_K h_K^2 \|\nabla(i_h p - i_h i_{2h} p)\|_P^2 \leqslant c_h \sum_{P\in\Omega_h} \alpha_K h_K^2 \|\nabla(p - i_{2h} p)\|_P^2.$$

Es bleibt gerade die Interpolation in den groben Raum $\tilde{L}_h = Q_{2h}$ von Grad r, bei $p \in H^r(\Omega)$ gilt:

$$S_h(i_h p, i_h p) \leqslant c_h \sum_{P\in\Omega_h} \alpha_K h_K^{2r} \|\nabla^r p\|_P^2.$$

Die Interpolation in den Raum der Polynome vom Grad r w"urde prinzipiell eine noch bessere Approximationsordnung erlauben. Zusammen mit den anderen Termen der Fehlerabsch"atzung aus Satz 3.25 zeigt sich jedoch so ein balanciertes Gesamtbild:

$$\begin{aligned}\|\mathbf{v} - \mathbf{v}_h\| + \|p - p_h\| &\leqslant c\Big(\|\nabla(\mathbf{v} - i_h\mathbf{v})\| + \|p - i_h p\| + S(i_h p, i_h p)^{\frac{1}{2}}\Big)\\ &\leqslant \Big(h^r\|\nabla^{r+1}\mathbf{v}\| + h^r\|\nabla^r p\| + \alpha h^r\|\nabla^r p\|\Big),\end{aligned}$$

mit $\alpha = \max\{\alpha_K\}$ und $h = \max\{h_K\}$.

b) Projektion in den Taylor-Hood Raum Eine sehr "ahnliche Konstruktion kann auf Basis der Taylor-Hood R"aume, also der Ans"atze $Q^r - Q^{r-1}$ erstellt werden. Wir w"ahlen f"ur $V_h \times L_h$ den den equal-order Raum $Q^r - Q^r$ und als stabilen Raum $V_h \times \tilde{L}_h$ das Taylor-Hood Element $Q^r - Q^{r-1}$. Der diskrete Projektionsoperator π_h kann wieder als Interpolation geschrieben werden. Zusammen mit dem Stabilisierungsterm (wie oben)

$$S_h(p_h, \xi_h) = \sum_{K\in\Omega_h} \alpha_K h_K^2 (\nabla(p_h - \pi_h p_h), \nabla(\xi_h - \pi_h \xi_h))_K,$$

ergeben sich wieder die drei Eigenschaften sowie eine optimale Approximationsordnung.

c) Alternativen f"ur den Fall $Q^1 - Q^1$ Zum Abschluss diskutieren wir nun noch einige Alternativen f"ur den einfachsten Fall von bilinearen equal-order Elementen $Q^1 - Q^1$ mit stabilen Ansatzraum $Q^1 - \text{iso}\, Q^1$. Der Projektionsoperator sei stets die Interpolation nach Ω_{2h}.

Eine erste Variante wurde bereits unter Punkt **a**) vorgestellt und definiert die Stabilisierungsform als

$$S_h(p_h, \xi_h) = \sum_{P \in \Omega_h} \alpha_K h_K^2 (\nabla(p_h - \pi_h p_h), \nabla(\xi_h - \pi_h \xi_h))_P.$$

Alternative passt in diesen Rahmen jedoch auch die urspr"ungliche Form der LPS-Methode, also die Projektion des Druckgradienten auf die patchweise Konstanten:

$$S_h(p_h, \xi_h) = \sum_{P \in \Omega_h} \alpha_K h_K^2 (\nabla p_h - \overline{\nabla p_h}, \nabla \xi_h - \overline{\nabla \xi_h})_P.$$

Bedingung (3.21) folgt wieder direkt aus der zugrundeliegenden Eigenschaft des Skalarprodukts. Es bleibt, den Stabilisierungsterm gegen"uber dem Projektionsfehler zu beschr"anken:

$$S_h(p_h, p_h) = \sum_P \alpha_K h_K^2 \|\nabla p_h - \overline{\nabla p_h}\|_P^2 \overset{!}{\geqslant} c \|p_h - \pi_h p_h\|^2.$$

Diese Beziehung kann wieder mit inversen Ungleichungen hergeleitet werden. Dabei ist insbesondere zu "uberpr"ufen, dass aus $p_h - \pi_h p_h = 0$ auch immer $\nabla p_h - \overline{\nabla p_h} = 0$ folgen muss.

Es bestehen weitere lokale M"oglichkeiten einen Stabilisierungsterm zu definieren. Hierzu sei Ω_P wieder eine Patch-Struktur auf Ω_h. Wir definieren den Stabilisierungsterm als

$$S_h(p_h, \xi_h) = \sum_{P \in \Omega_P} \alpha_P \sum_{e \in E_i(P)} h_e^3 \int_e [\partial_n p_h][\partial_n \xi_h] \, do,$$

wobei $E_i(P)$ die (vier) internen Kannten des Patches P sind und $[\cdot]$ der Sprung "uber die Normalableitung auf dem Rand ist. Auch hier k"onnen die Bedingungen mit lokalen Skalierungsargumenten mittels Transformation auf Referenzelement und mit Hilfe der Norm"aquivalenz nachgewiesen werden.

Anhand dieser Diskussion zeigt sich, dass die LPS-Methode eine ganze Klasse von verschiedenen Stabilisierungsverfahren ist. Die Wahl unterschiedlicher Stabilisierungsoperatoren hat Einfluss auf den Fehler der Approximation, ist jedoch auch in technischer Hinsicht relevant. Je nach Definition von Projektionen und Interpolationen f"uhrt die LPS-Methode zu Matrizen mit sehr gro"ser Zahl an Kopplungen.

3.4 Discretization of the Navier-Stokes equations

In the following we will discuss the Navier-Stokes equations

$$(\partial_t \mathbf{v}, \phi) + \frac{1}{Re}(\nabla \mathbf{v}, \nabla \phi) + ((\mathbf{v} \cdot \nabla)\mathbf{v}, \phi) - (p, \nabla \cdot \phi) = (\mathbf{f}, \phi) \quad \forall \phi \in H_0^1(\Omega; \Gamma_D)^d,$$
$$(\nabla \cdot \mathbf{v}, \xi) = 0 \qquad \forall \xi \in L_0^2(\Omega),$$

with $L_0^2(\Omega) := \{q \in L^2(\Omega) : \int_\Omega q dx = 0\}$ that includes the nonlinear *convection term* $((\mathbf{v} \cdot \nabla)\mathbf{v}, \phi)$. The *Reynolds number* Re appears and it will take an important role in determining the character of the equation. For $Re \to \infty$ the Navier-Stokes equations turn to the *Euler equations*

$$(Re \to \infty): \quad (\partial_t \mathbf{v}, \phi) + ((\mathbf{v} \cdot \nabla)\mathbf{v}, \phi) - (p, \nabla \cdot \phi) = (\mathbf{f}, \phi), \quad (\nabla \cdot \mathbf{v}, \xi) = 0,$$

a differential equation of first order that describes the flow of fluids where friction does not play a role. It is the dominant equation in aerodynamics of fast-flying planes at high altitude. On the other hand, the limit case $Re \to 0$ of the Navier-Stokes equations leads to the Stokes equations

$$(Re \to 0): \quad (\partial_t \mathbf{v}, \phi) + \frac{1}{Re}(\nabla \mathbf{v}, \nabla \phi) - (p, \nabla \cdot \phi) = (\mathbf{f}, \phi), \quad (\nabla \cdot \mathbf{v}, \xi) = 0,$$

as linear limit case for flows that are governed by friction. Numerically the Reynolds number gets important as the transport dominant case $Re \gg 1$ has different requirements than the friction dominant case $Re \ll 1$. Often, both regimes can act in one flow configuration. We must design numerical techniques that are able to deal with a problem of hyperbolic character (the Euler limit) and with elliptic character (the Stokes limit).

We start by discussing the numerical treatment of the nonlinearity for stationary equations.

3.4.1 Linearization of the Navier-Stokes equations

Discretizing the Navier-Stokes equations with finite elements leads to a quadratic nonlinear algebraic system of equations. Let $\mathbf{v}_h \in V_h$ be a finite element function with basis representation $\mathbf{v}_h := \sum_{i=1}^{N_h^v} \mathbf{v}_i \phi_h^i$. Then, the convective term gets

$$((\mathbf{v}_h \cdot \nabla)\mathbf{v}_h, \phi_h^i) = \sum_{k,l=1}^{N_h^v} \underbrace{((\phi_h^k \cdot \nabla)\phi_h^l, \phi_h^i)}_{=:\mathbf{N}_{ikl}} \mathbf{v}_k \mathbf{v}_l =: \mathbf{N}\mathbf{v}\mathbf{v}.$$

The discrete system is described by a tensor $\mathbf{N} = (\mathbf{N}_{ijk})_{ijk}$ of degree three. For reasons of efficiency it is often not feasible to assemble this tensor. Comparable to the computation of non-zero entries in the system matrix in Section 3.2.4 one could compute the number of non-zeros in $\mathbf{N}$. The memory (and computational) requirements are enormous.

Instead of direct discretization we first *linearize* the Navier-Stokes problem with fixed point iterations. Such a strategy was also considered in the existence proof for solutions to the Navier-Stokes equations in theorem 2.35. We will discuss various possibilities.

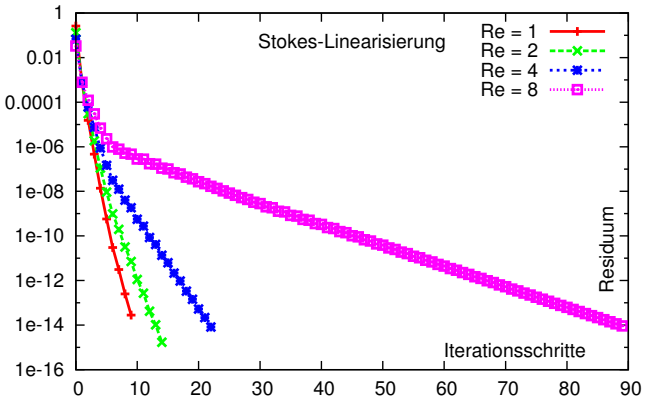


Figure 3.10: Convergenve of the Stokes linearization for increasing Reynolds numbers.

a) Stokes linearization The most simple possibility for a linearization is to consider the nonlinearity explicitly: Starting with an initial guess $\mathbf{v}^0$ we iteratively solve the following fixed point problem for $l = 1, 2, \ldots$

$$\frac{1}{Re}(\nabla \mathbf{v}^l, \nabla \phi) - (p^l, \nabla \cdot \phi) = (\mathbf{f}, \phi) - ((\mathbf{v}^{l-1} \cdot \nabla)\mathbf{v}^{l-1}, \phi)$$
$$(\nabla \cdot \mathbf{v}^l, \xi) = 0.$$

Every step consists of a Stokes problem and can be treated with help of the different techniques discussed before. Numerical experiments, however, show that such a simple iteration will only converge for very small Reynolds numbers $Re \ll 1$. Figure 3.10 shows the convergence of the Stokes linearization for the flow around an obstacle at different viscosities. Already the case $Re = 8$ requires 100 steps to solve the nonlinear problem with sufficient accuracy.

To analyze the convergence we subtract the exact solution $\{\mathbf{v}, p\}$ and diagonally test with the iteration error $\phi = \mathbf{v} - \mathbf{v}^l$ and $\xi := p - p^l$

$$\frac{1}{Re}\|\nabla(\mathbf{v} - \mathbf{v}^l)\|^2 = -\big(\mathbf{v} \cdot \nabla(\mathbf{v} - \mathbf{v}^{l-1}) + (\mathbf{v} - \mathbf{v}^{l-1}) \cdot \nabla \mathbf{v}^{l-1}, \mathbf{v} - \mathbf{v}^l\big).$$

We consider now the discrete case, i.e. $\mathbf{v} \in V_h$ such that all function spaces are finite dimensional such that we can use equivalence of norms. Then, it holds

$$\frac{1}{Re}\|\nabla(\mathbf{v} - \mathbf{v}^l)\|^2 \leqslant c_P\Big(\|\mathbf{v}\|_\infty \|\nabla(\mathbf{v} - \mathbf{v}^{l-1})\| + c_P\|\nabla \mathbf{v}^{l-1}\|_\infty \|\nabla(\mathbf{v} - \mathbf{v}^{l-1})\|\Big)\|\nabla(\mathbf{v} - \mathbf{v}^l)\|.$$

Hence

$$\|\nabla(\mathbf{v} - \mathbf{v}^l)\| \leqslant Re\, c_P\Big(\|\mathbf{v}\|_\infty + c_P\|\nabla \mathbf{v}^{l-1}\|_\infty\Big)\|\nabla(\mathbf{v} - \mathbf{v}^{l-1})\|,$$

which shows that convergence can only be guaranteed for small Reynolds numbers. Figure 3.10 shows that this analysis is too pessimistic, although rates are very low, we actually see convergence for a small range of numbers.

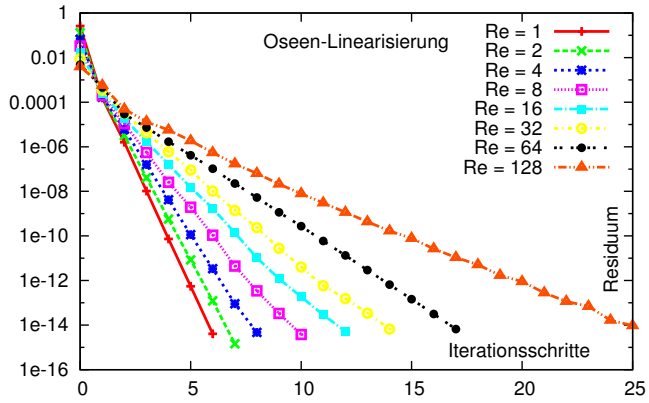


Figure 3.11: Convergence of the Oseen-Linearization for different Reynolds numbers.

It is important to note that this analysis is further inaccurate as we used equivalence of discrete norms. The constants can depend on the mesh parameter h such that the bounds are not robust for $h \to 0$.

b) Oseen-Linearization For larger Reynolds numbers we consider the *Oseen-Linearization*. We split the convective term into an explicit and into an implicit part. Starting with $\mathbf{v}^0$ we iterate for $l = 1, 2, \dots$ the fixed point problem

$$\frac{1}{Re}(\nabla \mathbf{v}^l, \nabla \phi) + ((\mathbf{v}^{l-1} \cdot \nabla)\mathbf{v}^l, \phi) - (p^l, \nabla \cdot \phi) = (\mathbf{f}, \phi)$$
$$(\nabla \cdot \mathbf{v}^l, \xi) = 0.$$

Every step asks for the solution of a (linear) *diffusion transport problem* that is called the *Oseen problem*. Figure 3.11 shows the convergence of the Oseen linearization. The required number of iteration once more increases with increasing Reynolds numbers. Compared to the Stokes linearization we are able to significantly increase the Reynolds number at much lower iteration counts. The iteration is fast for small Reynolds numbers $Re < 10$ such that less than 10 steps of the Oseen problem are required.

c) Newton-Linearization The most efficient method for solving the stationary Navier-Stokes equations is the Newton scheme. To derive it we start by repeating the Newton method in $\mathbb{R}^n$. Let $A : \mathbb{R}^n \to \mathbb{R}^n$ be a nonlinear map. We aim at finding $\mathbf{x} \in \mathbb{R}^n$ as solution to

$$A(\mathbf{x}) = b,$$

where $\mathbf{b} \in \mathbb{R}^n$ is a given right hand side.

Let $x^0 \in \mathbb{R}^n$ be an initial value. With $\mathbf{A}^l := \nabla A(\mathbf{x}^l) \in \mathbb{R}^{n \times n}$ we denote the Jacobian of A in $\mathbf{x}^l$. Then, the Newton scheme is given by the following iteration

$$\mathbf{A}^l \mathbf{w}^l = \mathbf{b} - A(\mathbf{x}^l), \quad \mathbf{x}^{l+1} = \mathbf{x}^l + \mathbf{w}^l, \quad l = 0, 1, 2, \dots \tag{3.22}$$

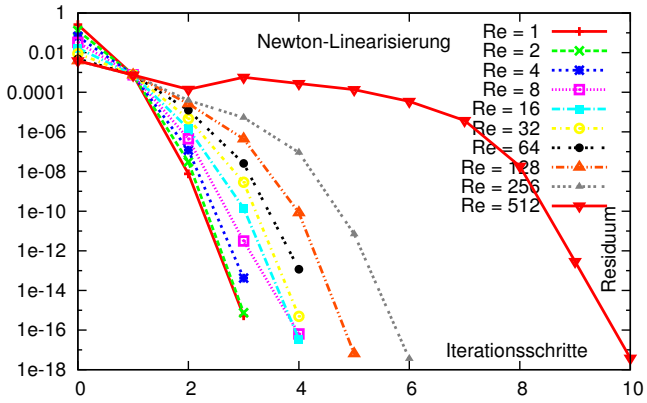


Figure 3.12: Convergence of the Newton-Linearization for different Reynolds numbers.

Every step asks for an inversion of the matrix $\mathbf{A}^l$. The expression $\mathbf{A}^l\mathbf{w}^l$ is the (scaled) directional derivative of A in $\mathbf{x}^l$ in direction $\mathbf{w}^l$. It holds

$$\mathbf{A}^l\mathbf{w}^l := \nabla A(\mathbf{x}^l)\mathbf{w}^l = \frac{d}{ds}A(\mathbf{x}^l + s\mathbf{w}^l)\Big|_{s=0}.$$

Hence every step of the Newton-Iteration (3.22) finds the optimal search direction $\mathbf{w}^l$

$$\frac{d}{ds}A(\mathbf{x}^l + s\mathbf{w}^l)\Big|_{s=0} = \mathbf{b} - A(\mathbf{x}^l),\ \mathbf{x}^{l+1} = \mathbf{x}^l + \mathbf{w}^l, \quad l = 0, 1, \ldots.$$

We transfer this notation to the solution of nonlinear partial differential equations. Let

$$U \in X: \quad A(U)(\Phi) = F(\Phi) \quad \forall \Phi \in X,$$

describe a PDE in variational formulation. By $A(\cdot)(\cdot)$ we define a form that is linear in the second argument. By $U^0 \in X$ we denote the initial value of the Newton iteration. Then we determine updates $W^l \in X$ by the problem

$$\frac{d}{ds}A(U^l + sW^l)(\Phi)\Big|_{s=0} = F(\Phi) - A(U^l)(\Phi), \quad U^{l+1} := U^l + W^l. \tag{3.23}$$

The directional derivative of $A(\cdot)(\cdot)$ at U^l in direction W^l is denoted by

$$A'(U^l)(W^l, \Phi) := \frac{d}{ds}A(U^l + sW^l)(\Phi)\Big|_{s=0}. \tag{3.24}$$

By $A'(\cdot)(\cdot,\cdot)$ a form is defined that might still be nonlinear in the first argument but that is linear in the second and third argument.

We specify the Newton scheme for the Navier-Stokes equations. It holds

$$A(U)(\Phi) = \frac{1}{Re}(\nabla\mathbf{v}, \nabla\phi) + ((\mathbf{v}\cdot\nabla)\mathbf{v}, \phi) - (p, \nabla\cdot\phi) + (\nabla\cdot\mathbf{v}, \xi), \quad F(\Phi) = (\mathbf{f}, \phi). \tag{3.25}$$

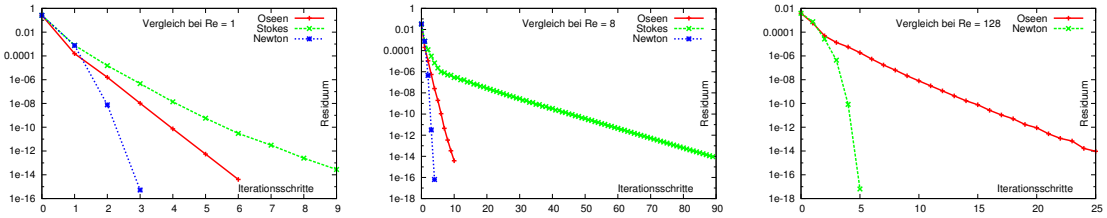


Figure 3.13: Comparison of the different linearization techniques for different Reynolds numbers (left $Re = 1$, middle $Re = 8$, right $Re = 128$). For $Re = 128$ we do not observe convergence of the Stokes linearization.

Let $U := \{\mathbf{v}, p\}$, $\Phi := \{\phi, \xi\}$ and $W := \{\mathbf{w}, q\}$. The directional derivative of the semilinear form $A(\cdot)(\cdot)$ is given as

$$A'(U)(W,\Phi) := \frac{d}{ds}A(U+sW)(\Phi)\Big|_{s=0} = \frac{1}{Re}(\nabla\mathbf{w},\nabla\phi)+((\mathbf{v}\cdot\nabla)\mathbf{w}+(\mathbf{w}\cdot\nabla)\mathbf{v},\phi)-(q,\nabla\cdot\phi)+(\nabla\cdot\mathbf{w},\xi).$$

Every step of the Newtons scheme (3.23) asks for the solution of a *linear* differential equation

$$\frac{1}{Re}(\nabla\mathbf{w},\nabla\phi)+((\mathbf{v}\cdot\nabla)\mathbf{w}+(\mathbf{w}\cdot\nabla)\mathbf{v},\phi)-(q,\nabla\cdot\phi) = (\mathbf{f},\phi)-\frac{1}{Re}(\nabla\mathbf{v},\nabla\phi)-((\mathbf{v}\cdot\nabla)\mathbf{v},\phi)+(p,\nabla\cdot\phi),$$
$$(\nabla\cdot\mathbf{w},\xi) = -(\nabla\cdot\mathbf{v},\xi).$$

This equation is a *diffusion-transport-reaction problem*. The zero-order term $((\mathbf{w}\cdot\nabla)\mathbf{v},\phi)$ is called a reaction term. We can discretize this problem with finite elements. Given sufficiently good initial values $U^0 := \{\mathbf{v}^0, p^0\}$ the Newton scheme will quadratically converge to a discrete solution $U = \{\mathbf{v}, p\}$. In Figure 3.12 we show the convergence history for the flow around an obstacle. Using the Newton scheme we observe very low iteration counts for a wide range of Reynolds numbers. Only the case $Re = 512$ shows a significant raise in iteration numbers. The reason for this is found in the solution properties of the Navier-Stokes equations. We have seen that the stationary problem only admits a solution for small Reynolds numbers. Here, for $Re = 512$ there does not exist a physically relevant stationary solution. At $Re = 300 \sim 400$ there is a transition to a nonstationary flow pattern. As there is no stable stationary solution the Newton scheme cannot converge. In all other cases we can clearly observe quadratic convergence. Every step yields a doubling of the accurate digits. Finally we compare all techniques in Figure 3.13.

The Newton scheme is the most powerful technique for nonlinear partial differential equations and in particular for the Navier-Stokes problem. The analysis of the linear problems is difficult as we cannot control the sign (positivity) of the reaction term

$$((\mathbf{w}\cdot\nabla)\mathbf{v},\phi).$$

The matrix $\nabla\mathbf{v}$ has both positive and negative eigenvalues which follows from $\mathrm{tr}\,(\nabla\mathbf{v}) = \mathrm{div}\,\mathbf{v} = 0$. Existence of solutions and convergence of the scheme can only be shown for small problem data.

A convergence analysis of the finite element discretization of the nonlinear Navier-Stokes problem is not easily possible. We consider the Oseen problem.

Lemma 3.26 (Finite element convergence of the Oseen problem)**.** Let $\mathbf{w} \in V_0 \subset H_0^1(\Omega)^d$ be a divergence free transport field. Let $\mathbf{v}_h \in V_h$ and $p_h \in L_h$ be the solution in an inf-sup stable finite element pair of

$$\nu(\nabla \mathbf{v}_h, \nabla \phi_h) + ((\mathbf{w} \cdot \nabla)\mathbf{v}_h, \phi_h) - (p_h, \nabla \cdot \phi_h) + (\nabla \cdot \mathbf{v}_h, \xi_h) = (\mathbf{f}, \phi_h)$$

for all $\phi_h \in V_h$ and $\xi_h \in L_h$. It holds

$$\|\nabla(\mathbf{v} - \mathbf{v}_h)\| + \|p - p_h\| \leqslant c(\gamma_h, \nabla \mathbf{v})\Big(\inf_{\phi_h \in V_h} \|\nabla(\mathbf{v} - \phi_h)\| + \inf_{\xi_h \in L_h} \|p - \xi_h\|\Big),$$

with the constant $\gamma_h \geqslant \gamma$ of the inf-sup condition.

Beweis: (*i*) For the error $\mathbf{v} - \mathbf{v}_h$ and $p - p_h$ we get the following Galerkin orthogonality

$$(\nabla \cdot (\mathbf{v} - \mathbf{v}_h), \xi_h) = 0, \quad \nu(\nabla(\mathbf{v} - \mathbf{v}_h), \nabla \phi_h) + (\mathbf{v} \cdot \nabla(\mathbf{v} - \mathbf{v}_h), \phi_h) = (p - p_h, \nabla \cdot \phi_h) = 0.$$

(*ii*) For $\mathbf{e}_h^v := \mathbf{v} - \mathbf{v}_h$ and $e_h^p := p - p_h$ it holds with arbitrary $\phi_h \in V_h$

$$\nu \|\nabla \mathbf{e}_h^v\|^2 = \nu(\nabla \mathbf{e}_h^v, \nabla(\mathbf{v} - \phi_h)) + \nu(\nabla \mathbf{e}_h^v, \nabla(\phi_h - \mathbf{v}_h)).$$

The first term can be bound with help of an interpolation estimate. For the last term we use Galerkin orthogonality

$$\begin{aligned}
\nu(\nabla \mathbf{e}_h^v, \nabla(\phi_h - \mathbf{v}_h)) &= (e_h^p, \nabla \cdot (\phi_h - \mathbf{v}_h)) - (\mathbf{v} \cdot \nabla \mathbf{e}_h^v, \phi_h - \mathbf{v}_h) \\
&= (p - \xi_h, \nabla \cdot \mathbf{e}_h^v) + (e_h^p, \nabla \cdot (\phi_h - \mathbf{v})) - \underbrace{(\mathbf{v} \cdot \nabla \mathbf{e}_h^v, \mathbf{e}_h^v)}_{=0} + (\mathbf{v} \cdot \nabla \mathbf{e}_h^v, \mathbf{v} - \phi_h) \\
&\leqslant \|p - \xi_h\| \, \|\nabla \mathbf{e}_h^v\| + \|e_h^p\| \, \|\nabla(\mathbf{v} - \phi_h)\| + c\|\nabla \mathbf{v}\| \, \|\nabla \mathbf{e}_h^v\| \, \|\nabla(\mathbf{v} - \phi_h)\|.
\end{aligned}$$

Together it holds

$$\frac{\nu}{2}\|\nabla \mathbf{e}_h^v\|^2 \leqslant \Big(\nu^{-1} + \nu^{-1} c^2 \|\nabla \mathbf{v}\|^2 + \frac{1}{4\epsilon}\Big) \inf_{\phi_h \in V_h} \|\nabla(\mathbf{v} - \phi_h)\|^2 + \inf_{\xi \in L_h} \|p - \xi_h\| + \epsilon \|e_h^p\|.$$

(*iii*) The pressure is estimated with the discrete inf-sup condition as

$$\begin{aligned}
\|p - p_h\| &\leqslant \|p - \xi_h\| + \|p_h - \xi_h\| \leqslant \|p - \xi_h\| + \gamma_h^{-1} \sup_{\phi_h \in V_h} \frac{(\xi_h - p_h, \nabla \cdot \phi_h)}{\|\nabla \phi_h\|} \\
&= \|p - \xi_h\| + \gamma_h^{-1} \sup_{\phi_h \in V_h} \frac{(e_h^p, \nabla \cdot \phi_h)}{\|\nabla \phi_h\|} + \gamma_h^{-1} \sup_{\phi_h \in V_h} \frac{(p - \xi_h, \nabla \cdot \phi_h)}{\|\nabla \phi_h\|} \\
&\leqslant \|p - \xi_h\| + \gamma_h^{-1} \sup_{\phi_h \in V_h} \frac{(\nabla \mathbf{e}_h^v, \nabla \phi_h) + (\mathbf{v} \cdot \nabla \mathbf{e}_h^v, \phi_h)}{\|\nabla \phi_h\|} + \gamma_h^{-1} \|p - \xi_h\|.
\end{aligned}$$

Hereby we get

$$\|p - p_h\| \leqslant (1 + \gamma_h^{-1}) \inf_{\xi_h \in Q_h} \|p - \xi_h\| + \gamma_h^{-1}(1 + \|\nabla \mathbf{v}\|)\|\nabla \mathbf{e}_h^v\|.$$

(iv) Finally, both estimates are combined by taking

$$\epsilon = \frac{\nu \gamma_h}{4(1 + \|\mathbf{v}\|)}$$

and we obtain the error bound. □

3.5 Discretization of the time-dependent Navier-Stokes equations

The stationary Navier-Stokes equations in the case of Dirichlet boundary conditions have for every Reynolds number a solution (see Theorem 2.35). This solution is unique only if the Reynolds number is very small:

$$Re \lesssim 1.$$

In practice, problems however deal with high Reynolds numbers and an instationary behavior appears. Thus, we have to consider the time-dependent Navier-Stokes equations

$$\partial_t \mathbf{v} + (\mathbf{v} \cdot \nabla)\mathbf{v} - \nu\Delta\mathbf{v} + \nabla p = \mathbf{f}, \quad \nabla \cdot \mathbf{v} = 0, \quad \mathbf{v}\Big|_{t=0} = \mathbf{v}^0, \quad \mathbf{v}\Big|_{\partial\Omega} = \mathbf{g}. \tag{3.26}$$

We can consider these equations as a initial-boundary value problem. The transistion to a non-stationary flow does not appear at a fixed Reynolds number. Depending on the flow configuration this transition can happen in the area of $Re \approx 50 \sim 200$ (like in the case of a flow around an obstacle, see Figures 1.15 to 1.17) or at higher Reynolds numbers, for example $Re \approx 10000$ as in the case of driven-cavity.

Even in the case of very small Reynolds numbers, a non-stationary flow behavior is possible, if this is due to non-stationary data, like time-dependent boundary values $\mathbf{g}(x, t)$. Such non-stationarities, however, have to be distinguished from the *inherent non-stationarities* which appear despite stationary data. A typical pattern of these non-stationarities is the *von-Karman vortex street* behind an obstacle, see e.g. Figure 1.16. An accurate approximation of the dynamics of problems with inherent non-stationarities is a far more difficult challenge that the simulation of problems with a non-stationary behavior due to time-dependent data.

For the time discretization of the time-dependent Navier-Stokes equations, we can discuss approaches which we know from parabolic equations. The *Rothe-method* considers first a time discretization, then a space discretization. The momentum equation is discretized with a classical time step method. The incompressibility constraint is postulated as a constraint in each time step. For example, the discretization with an implicit Euler scheme on a fixed time grid $t_0 < t_1 < \cdots < t_M$ with $k_m := t_m - t_{m-1}$ leads to a sequence of quasi-stationary problems

$$\mathbf{v}_k^m + k_m(\mathbf{v}_k^m \cdot \nabla)\mathbf{v}_k^m - k_m\nu\Delta\mathbf{v}_k^m + k_m\nabla p_k^m = \mathbf{v}_k^{m-1} + k_m\mathbf{f}, \quad \nabla \cdot \mathbf{v}^m = 0, \quad m = 1, \ldots, M,$$

which are then discretized in space. On the other hand, the *method of lines* first introduces a discretization in space (e.g. using finite elements) leading to a system of ordinary differential equations for a space discrete solution $\{\mathbf{v}_h, p_h\} : [0, T] \to \mathbb{R}^N$

$$\frac{d}{dt}\mathbf{v}_h + N_h(\mathbf{v}_h, \mathbf{v}_h) + \nu A_h\mathbf{v}_h + B_h\mathbf{v}_h = \mathbf{f}_h, \quad C_h\mathbf{v}_h = 0, \quad t \geqslant 0.$$

Here, A_h, B_h and C_h describe matrices and N_h is a tensor of third kind. This space discrete equation is also called a *differential-algebraic system* (DAE) of order 2. The *algebraic* constraint

$C_h \mathbf{v}_h = 0$ defines the manifold W_h in the space $V_h \subset H_0^1(\Omega)^d$, on which the dynamics are happening. A discretization in time can then be done with a usual time stepping scheme.

A third possibility to discretize the time-dependent Navier-Stokes equations is to discretize space and time simultaneously. For this, we consider the Navier-Stokes equations in a variational form and utilize a simultaneous Galerkin discretization in space and time. The resulting schemes are sometimes algebraic equvalent to corresponding time-stepping schemes. An advantage of a simultaneous space-time discretization is the mathematical form which allows to consider residual-based error analysis and a posteriori error estimation.

Finally, a fourth option is a so-called *projection approach*. Here, the momentum equation is separated from the incompressibility in the time approximation. By this *operator-splitting* approach it can be avoided to solve a coupled saddle point problem in each iteration.

3.5.1 The Rothe-methode for the discretization of the Navier-Stokes equations

For the temporal discretization, we introduce a time grid in the interval $I = [0, T]$ by:

$$0 = t_0 < t_1 < \cdots < t_M = T, \quad k_m := t_m - t_{m-1}, \quad k := \sup_m k_m.$$

The solution $\{\mathbf{v}(x,t), p(x,t)\}$ is approximated by the time-discrete solutions $\{\mathbf{v}_k^m(x), p_k^m(x)\}$ for $m = 0, \ldots, M$. As a basis for various methods, we consider the *single-step-θ-scheme*:

$$\begin{aligned} &\mathbf{v}_k^m + k_m\theta(\mathbf{v}_k^m \cdot \nabla)\mathbf{v}_k^m - k_m\theta\nu\Delta\mathbf{v}_k^m + k_m\nabla p_k^m = \mathbf{v}_k^{m-1} + k_m\theta\mathbf{f}^m + k_m(1-\theta)\mathbf{f}^{m-1} \\ &- k_m(1-\theta)\mathbf{v}_k^{m-1} \cdot \nabla\mathbf{v}_k^{m-1} + k_m(1-\theta)\Delta\mathbf{v}_k^{m-1}, \quad \nabla \cdot \mathbf{v}_k^m = 0, \qquad m = 1, \ldots, M, \end{aligned} \tag{3.27}$$

where we denote with $\mathbf{f}^m := \mathbf{f}(t_m)$ the right-hand side at the time point t_m. The incompressibility is postulated at each time point t_m, i.e. independent of the $\theta \in [0,1]$, implicitly.

In the case $\theta = 0$, the θ-scheme is the *explicit Euler method*. Due to the saddle point structure and the implicit treatment of the incompressibility constraint, this method applied to the Navier-Stokes equations is not an explicit scheme in the proper sense:

$$\mathbf{v}_k^m + k_m\nabla p_k^m = \mathbf{v}_k^{m-1} + k_m\mathbf{f}^{m-1} - k_m(\mathbf{v}_k^{m-1} \cdot \nabla)\mathbf{v}_k^{m-1} + k_m\Delta\mathbf{v}_k^{m-1}, \quad \nabla \cdot \mathbf{v}_k^m = 0. \tag{3.28}$$

The problem can then be discretized in space using e.g. finite element methods. As in the case of parabolic equations, the explicit Euler method converges linearly in time $O(k)$. If we choose only first order accurate time step methods, the discretization error is very unbalanced and we need very smal time steps. In particular, the explicit Euler method requires a restrictive step size condition:

$$k \leqslant \nu^{-1}h^2.$$

For $\theta = 1$, we obtain the *implicit Euler method*:

$$\mathbf{v}_k^m + k_m(\mathbf{v}_k^m \cdot \nabla)\mathbf{v}_k^m - k_m \nu \Delta \mathbf{v}_k^m + k_m \nabla p_k^m = \mathbf{v}_k^{m-1} + k_m \mathbf{f}^m, \quad \nabla \cdot \mathbf{v}_k^m = 0. \quad (3.29)$$

This method converges independent of the time step size linearly $O(k)$.

Of particular importance is the *Crank-Nicolson method* in the case of $\theta = \frac{1}{2}$:

$$\mathbf{v}_k^m + \frac{k_m}{2}(\mathbf{v}_k^m \cdot \nabla)\mathbf{v}_k^m - \frac{k_m}{2}\nu \Delta \mathbf{v}_k^m + k_m \nabla p_k^m = \mathbf{v}_k^{m-1} + \frac{k_m}{2}\mathbf{f}^m + \frac{k_m}{2}\mathbf{f}^{m-1}$$
$$- \frac{k_m}{2}(\mathbf{v}_k^{m-1} \cdot \nabla)\mathbf{v}_k^{m-1} + \frac{k_m}{2}\Delta \mathbf{v}_k^{m-1}, \quad \nabla \cdot \mathbf{v}_k^m = 0, \qquad m = 1, \dots, M. \quad (3.30)$$

The Crank-Nicolson method converges quadratically $O(k^2)$.

The analysis of solvability and approximation properties for the time-discrete Navier-Stokes equations is very involved. In particular, aspects like handling different meshes at each time point, regularity of a solution, smoothing of irregular initial conditions require extensive investigations. We refer to the literature here, e.g. [19, 18, 17, 16].

Stability of time stepping schemes In order to discuss the time stepping schemes, we not only have to consider the approximation order $O(k^\alpha)$ regarding the time step size but also especially the stability aspects. Here, two questions are important:

- If we have small disturbances, e.g. by inaccuracy of the data or by numerical rounding errors, do they influence the global solution behavior? We would expect from a proper method that such disturbances decay (exponentially).
- Is the method able to preserve the energy of the flow? For example, small vortices shall be approximated correctly and not be damped by numerical dissipation.

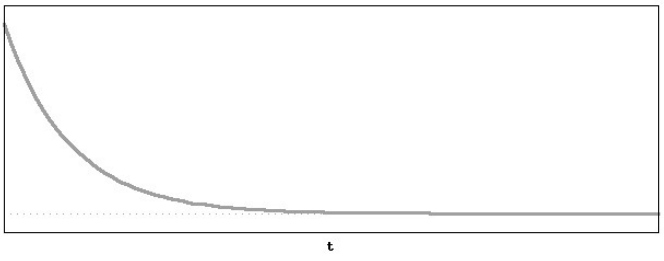
We investigate stability properties considering an easy scalar test equation (ODE):

$$u(0) = 1, \quad u'(t) = \lambda u(t), \quad t \geqslant 0, \ \lambda \in \mathbb{C}.$$

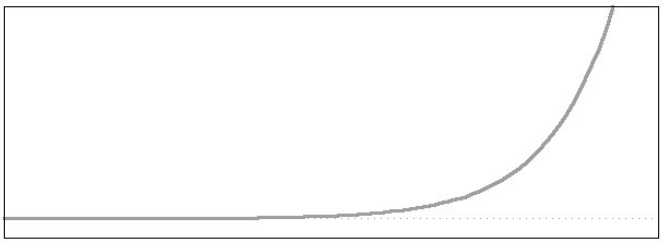
The solution can be described analytically as

$$u(t) = e^{\lambda t}.$$

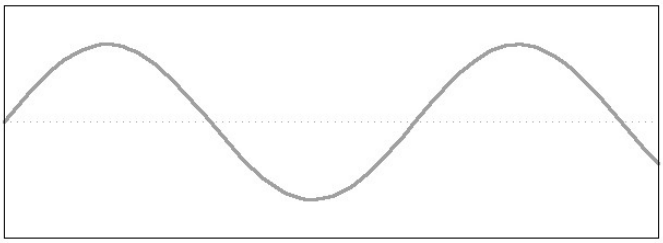
For $\lambda \in \mathbb{C}$ with real part $Re(\lambda) < 0$, the solution decays exponentially:



For values $\lambda \in \mathbb{C}$ with positive real part, the solution increases exponentially:



For purely imaginary $\lambda = i$ the solution is a wave



Each step of the θ-scheme applied to this model problem is given by:

$$u^m - \theta k\lambda u^m = u^{m-1} + (1-\theta)k\lambda u^{m-1} \quad \Leftrightarrow \quad u^m = \underbrace{\frac{1+(1-\theta)k\lambda}{1-\theta k\lambda}}_{=R(\lambda k)} u^{m-1}.$$

With the amplification factor $R(\lambda k)$ we can shortly write

$$u^m = R(\lambda k)u^{m-1}.$$

With every consistent method (of order $O(k^r)$) this factor is necessarily an approximation of the exponential function:

$$u(t_m) = R(\lambda k)u(t_m - k) + O(k^r) \;\Leftrightarrow\; e^{\lambda t_m} = R(\lambda k)e^{\lambda(t_m-k)} + O(k^r) \;\Leftrightarrow\; R(\lambda k) = e^{\lambda k} + O(k^r).$$

In the following, we set $z := \lambda k$. In the case $Re(\lambda) < 0$, the solution grows exponentially. In the case $Re(\lambda) < 0$ the solution decays exponentially to zero. In this case, we expect a corresponding behavior for the discrete solution. For the factor $R(\lambda k)$, we have:

$$R(z) = \frac{1+(1-\theta)z}{1-\theta z} = 1 + z + \theta z^2 + O(|z|^3) = e^z + \left(\theta - \frac{1}{2}\right) z^2 + O(|z|^3). \tag{3.31}$$

For $\theta = \frac{1}{2}$, the approximation is quadratic, otherwise it is linear. This is in accordance to our previous results for the implicit and explicit Euler method and the Crank-Nicolson method. We distinguish for $Re(\lambda) < 0$ the following stability terms:

A-stability holds, if

$$|R(\lambda)| \leqslant 1$$

for $Re(\lambda) \leqslant 0$. It follows local convergence of the solution for $t \in [0, T]$ for a $T < \infty$.

Global stability holds, if

$$\limsup |R(\lambda)| \leqslant 1 - O(k) \quad (Re(\lambda) \to -\infty)$$

It follows global convergence for $t \geqslant 0$.

Strong A-stability holds, if

$$\limsup |R(\lambda)| \leqslant 1 - \delta < 1 \quad (Re(\lambda) \to -\infty)$$

with a $\delta > 0$. From this, the damping property of numerical methods follow. Numerical errors are damped.

In the following, we discuss the important time stepping scheme with regards to their stability properties:

The explicit Euler method $\theta = 0$ It holds

$$R(z) = 1 + z.$$

This method fulfills none of the stability terms since

$$|R(z)| \to \infty \quad (Re(z) \to -\infty).$$

Further, for $Re(z) \to 0$ it holds:

$$|R(i)| = |1 + i| = \sqrt{2}.$$

Natural vibrations are even amplified. Due to the missing stability, the explicit Euler method is not suitable for the discretization of the Navier-Stokes equations.

The implicit Euler method $\theta = 1$ For the implicit Euler method, the amplification factor is given by

$$R(z) = \frac{1}{1 - z}$$

And it holds for $Re(z) < 0$:

$$|R(z)| \leqslant \frac{1}{1 + |Re(z)|} \to 0 \quad (Re(\lambda) \to -\infty).$$

The implicit Euler method is strong A-stable and global stable. For purely imaginary parts, it holds:

$$|R(i)| = \frac{1}{|1 - i|} = \sqrt{\frac{1}{2}},$$

and natural vibrations are strongly damped. Arbitrary large time steps can be used.

The method is well suited in order to approximate stationary solutions. Often, it is not possible to obtain these in a stationary solution process. Bad conditioning of the problem or strong nonlinearity prevent convergence of a solution method. Then, the solution can be approximated via "pseudo-time stepping schemes". The additional mass term $k^{-1}(v^m, \phi)_\Omega$ provides denfiniteness and reduces (for small k) the influence of the nonlinearity. The implicit Euler scheme is because of the large dissipation not suitable for the simulation of non-stationary flows. The transition to the non-stationarity follows often only for unphysically large Reynolds numbers and using very small time steps.

The Crank-Nicolson method $\theta = \frac{1}{2}$ In contrast to the Euler method, the Crank-Nicolson method is a method of second order:

$$R(z) = e^z + O(|z|^3).$$

The amplification factor is given by

$$R(z) = \frac{1 + \frac{z}{2}}{1 - \frac{z}{2}} = \frac{2 + z}{2 - z}$$

For $z = z_r + iz_i$ with negative real part $z_r < 0$ it holds

$$\left|1 + \frac{z}{2}\right| = \sqrt{\left(1 + \frac{z_r}{2}\right)^2 + \frac{z_i^2}{4}} \leqslant \sqrt{\left(1 - \frac{z_r}{2}\right)^2 + \frac{z_i^2}{4}} = \left|1 - \frac{z}{2}\right|,$$

thus the method is A-stable. For $Re(z) \to -\infty$ the amplification factor fulfills:

$$|R(z)| = \left|\frac{1 + \frac{z}{2}}{1 - \frac{z}{2}}\right| = \left|\frac{\frac{2}{z} + 1}{\frac{2}{z} - 1}\right| \to 1 \quad (Re(z) \to -\infty).$$

This method is not globally stable and not strong A-stable. For $z = i$ it holds:

$$|R(i)| = 1,$$

thus the Crank-Nicolson method is exact energy preserving. In practice, this method is used very often because of the second order. Disturbances (e.g. through rounding errors or in the initial data) are, however, not sufficiently damped. As a solution, the Crank-Nicolson scheme can be combined with e.g. the implicit Euler scheme. For this, in the first few (e.g. 2-5) steps, the implicit Euler schemes is used before the Crank-Nicolson method is then applied. By this, the errors in the initial data are damped and eventually the good convergence of the Crank-Nicolson method is exploited. For numerical long-time simulations, further implicit Euler steps can be used at fixed time points $t_0 + T_0$, $t_0 + 2T_0$, $t_0 + 3T_0, \ldots$ in order to damp possible rounding effects. This method is also known as *Rannacher- time stepping* and often used in finance-mathematical applications.

Das implizit-geshiftete Crank-Nicolson-Verfahren $\theta = \frac{1}{2}+k$ f"ur $k \ll 1$ Eine andere M"oglichkeit zum Erlangen von besserer Stabilit"at des Crank-Nicolson-Verfahrens ist ein *impliziter Shift* durch Wahl von $\theta = \frac{1}{2} + k$. Dieses Verfahren ist trotz Shift immer noch von zweiter Ordnung in k. Es gilt mit (3.31) und $z = \lambda k$

$$R(\lambda k) = e^{\lambda k} + k(k\lambda)^2 + O(|k|^3).$$

Halten wir k fest, so gilt f"ur $Re(z) \to -\infty$:

$$|R(z)| = \left|\frac{1 + \left(\frac{1}{2} - k\right) z}{1 - \left(\frac{1}{2} + k\right) z}\right| = \left|\frac{2z^{-1} + (1 - 2k)}{2z^{-1} - (1 + 2k)}\right| \xrightarrow[Re(z)\to-\infty]{} \left|\frac{1 - 2k}{1 + 2k}\right| = 1 - 4k + O(k^2).$$

Dieses Verfahren ist also *global stabil*, allerdings nicht *stark A-stabil*. In der Praxis erweist sich das geshiftete-Crank-Nicolson Verfahren als gen"ugend robust um zu verhindern, dass im Laufe der Zeitschritte angeh"aufte Rundungsfehler dominant werden. Die Stabilit"at des Crank-Nicolson-Verfahrens mit Shift reicht hingegen nicht aus um etwa fehlende Regularit"at in den Anfangsdaten hinreichend zu gl"atten. F"ur rein imagin"are Anteile gilt:

$$|R(i)| = \left|\frac{1 + \left(\frac{1}{2} - k\right) i}{1 - \left(\frac{1}{2} + k\right) i}\right| = \left(\frac{1 + \left(\frac{1}{2} - k\right)^2}{1 + \left(\frac{1}{2} + k\right)^2}\right)^{\frac{1}{2}} = 1 - O(k).$$

F"ur kleine k ist das Verfahren somit gut energieerhaltend.

Das Teilschritt-θ-Verfahren Das sogenannte *fractional-step-θ-scheme* kombiniert drei aufeinander folgende Schritte des Einschritt-θ-Verfahrens um m"oglichst optimale Eigenschaften eines Gesamtverfahrens zu erreichen:

$$t_{m-1} \xrightarrow[\alpha_1 k_m]{\theta_1} t_{m-1+\alpha_1} \xrightarrow[\alpha_2 k_m]{\theta_2} t_{m-\alpha_3} \xrightarrow[\alpha_3 k_m]{\theta_3} t_m.$$

F"ur jeden der drei Teilschritte kann ein eigener Wert f"ur θ sowie f"ur die Teilschrittweite αk gew"ahlt werden. Dabei gilt: $\theta_i \in (0, 1]$ und $\alpha_i > 0$ sowie $\alpha_1 + \alpha_2 + \alpha_3 = 1$. Es stehen nun 6 verschiedene Parameter zur freien Wahl um ein konkretes Verfahren zu erreichen. Um diese Anzahl der freien Parameter zu reduzieren betrachten wir zun"achst nur symmetrische Verfahren mit $\alpha_1 = \alpha_3 =: \alpha$ und somit $\alpha_2 = 1 - \alpha_1 - \alpha_3 = 1 - 2\alpha$. Ebenso w"ahlen wir $\theta_1 = \theta_3 =: \theta$ und $\theta_2 = 1 - \theta$. Der Verst"arkungsfaktor des so gewonnen Verfahrens berechnet sich als:

$$R(z) = \left(\frac{1 + (1 - \theta)\alpha z}{1 - \theta\alpha z}\right)^2 \frac{1 + \theta(1 - 2\alpha)z}{1 - (1 - \theta)(1 - 2\alpha)z}.$$

Zur Approximation der Exponentialfunktion hat der Verst"arkungsfaktor die Reihenentwicklung:

$$R(z) - \exp(z) = (1 - 2\theta)\left(\frac{1}{2} - 2\alpha + \alpha^2\right) z^2 + O(|z|^3).$$

Quadratische Konvergenz erhalten wir f"ur den Wert $\alpha = 1 - \sqrt{\frac{1}{2}} \approx 0.2929$ bei beliebiger Wahl von θ. Alternativ kann $\theta = \frac{1}{2}$ und α beliebig gew"ahlt werden. Dieses Verfahren entspricht jedoch der Hintereinanderreihung von drei Schritten des Crank-Nicolson Verfahrens, ist also nicht weiter Interessant.

Betrachten wir den Grenzwert $\mathrm{Re}(z) \to -\infty$ so gilt f"ur den Verst"arkungsfaktor

$$|R(z)| \to \frac{1-\theta}{\theta}.$$

F"ur beliebige $\theta \in (0, 1]$ ist das Teilschritt-θ-Verfahren stark A-stabil.

Zur kompakten Schreibweise des Verfahren angewendet auf die Navier-Stokes Gleichungen f"uhren wir die folgende Operatorschreibweise ein:

$$\mathcal{A}(\mathbf{v}) := \mathbf{v} \cdot \nabla - \nu\Delta, \quad \mathcal{B} := \nabla, \quad -\mathcal{B}^* := \operatorname{div}.$$

Dann nimmt das Teilschritt-θ-Verfahren die folgende Form an:

$$\begin{aligned}
\mathbf{v}^{m-1+\theta} + \alpha\theta k\mathcal{A}(\mathbf{v}^{m-1+\theta})\mathbf{v}^{m-1+\theta} + k\mathcal{B}p^{m-1+\theta} &= \mathbf{v}^{m-1} - \alpha(1-\theta)k\mathcal{A}(\mathbf{v}^{m-1})\mathbf{v}^{m-1} \\
\mathbf{v}^{m-\theta} + (1-2\alpha)(1-\theta)k\mathcal{A}(\mathbf{v}^{m-\theta})\mathbf{v}^{m-\theta} + k\mathcal{B}p^{m-\theta} &= \mathbf{v}^{m-1+\theta} - (1-2\alpha)\theta k\mathcal{A}(\mathbf{v}^{m-1+\theta})\mathbf{v}^{m-1+\theta} \\
\mathbf{v}^{m} + \alpha\theta k\mathcal{A}(\mathbf{v}^{m})\mathbf{v}^{m} + k\mathcal{B}p^{m} &= \mathbf{v}^{m-\theta} - \alpha(1-\theta)k\mathcal{A}(\mathbf{v}^{m-\theta})\mathbf{v}^{m-\theta}
\end{aligned}$$

Eine Vereinfachung des Verfahrens kann durch die Wahl

$$\alpha\theta = (1-2\alpha)(1-\theta) \quad \Leftrightarrow \quad \theta = \frac{1-2\alpha}{1-\alpha} \approx 0.5858,$$

erreicht werden. Durch diese Wahl ist der Vorfaktor im impliziten Teil des Operators stets gleich:

$$\mathcal{L}(\mathbf{v})[\mathbf{v}, p] := [\mathrm{id} + \alpha\theta k\mathcal{A}(\mathbf{v})]\mathbf{v} + k\mathcal{B}p.$$

Angenommen, der Operator $\mathcal{A}$ w"are linear, so k"onnte jeder Teilschritt des Verfahrens mit ein und derselben Systemmatrix berechnet werden. Aufgrund des sehr aufw"andigen Matrix-Aufbaus bei der Finite Elemente Methode spielt diese "Uberlegung eine Rolle. Da die Navier-Stokes Gleichungen jedoch nichtlinear sind und die Systemmatrix sowieso von der aktuellen Approximation abh"angt ist diese "Uberlegung bei den Navier-Stokes Gleichungen nicht wesentlich. Die spezielle Wahl $\theta = (1-2\alpha)/(1-\alpha)$ hat sich dennoch etabliert. Abschlie"send betrachten wir f"ur diesen speziellen Wert von θ noch die Erhaltungseigenschaft bei rein imagin"aren Anteilen $z = \mathrm{i}$. Es gilt mit

$$R\left(\frac{\mathrm{i}}{10}\right) \approx 0.99999996, \quad R(\mathrm{i}) \approx 0.999696, \quad R(10\mathrm{i}) \approx 0.828,$$

eine fast optimale Erhaltung von freien Schwingungen im Fall kleiner und moderater Schrittweiten. Mit quadratischer Konvergenzordnung, starker A-Stabilit"at und sehr geringer Dissipativit"at ist das Teilschritt-θ-Verfahren ein nahezu optimales Verfahren zur Zeitdiskretisierung bei den Navier-Stokes Gleichungen.

Numerischer Vergleich Wir betrachten das System von Anfangswertaufgaben:

$$u'(t) + Au(t) = 0, \quad u(0) = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \quad A := \begin{bmatrix} 0 & 10^\gamma & -1 & 0 \\ 0 & 10^\gamma & 0 & 0 \\ 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Die Matrix hat die Eigenwerte

$$\lambda_1 = 10^\gamma,\ \lambda_2 = i,\ \lambda_3 = -i,\ \lambda_4 = 1,$$

das System ist also bei $\gamma > 1$ sehr steif. Die L"osung ist gegeben durch

$$u(t) = \begin{pmatrix} e^{-10^\gamma t} + \sin(t) \\ e^{-10^\gamma t} \\ \cos(t) \\ e^{-t} \end{pmatrix}.$$

Die L"osung enth"alt sowohl Komponenten, welche sehr schnell abklingen als auch Schwingungen, sowie Kombinationen davon. Anhand dieser Aufgabe lassen sich also s"amtliche Stabilit"atsbegriffe und Eigenschaften "uberpr"ufen. In Abbildung 3.14 zeigen wir die L"osung der ersten L"osungskomponente bei Diskretisierung mit dem impliziten Euler-Verfahren, dem Crank-Nicolson als auch dem Fractional-Step-Theta-Verfahren. Das explizite Euler-Verfahren divergiert aufgrund der Verst"arkung von Schwingungen unmittelbar.

In Abbildung 3.15 stellen wir die dritte L"osungskomponente $u_k^3(t)$ sowie den Fehler $u^3(t) - u_k^3(t)$ f"ur das explizite, sowie das implizite Euler-Verfahren, f"ur das Crank-Nicolson und f"ur das Fractional-Step-Theta-Verfahren dar. Die dritte Komponente geh"ort besitzt einen rein imagin"aren Anteil. Die wesentliche Eigenschaft ist also die Energieerhaltung. Wie erwartet d"ampft der implizite Euler zu sehr, der explizite verst"arkt die Schwingung. Sowohl Crank-Nicolson, als auch Fractional-Step-Theta stellen den L"osungsverlauf sehr gut dar.

3.5.2 Projektionsmethoden

Die sogenannten Projektionsverfahren verfolgen einen *Operator-Splitting Ansatz*: In jedem Zeitschritt $t_{m-1} \to t_m$ wird zun"achst ein neues Geschwindigkeitsfeld $\tilde{v}^m$ berechnet, ohne die Divergenzfreiheit zu beachten. In einem zweiten Schritt wird diese vorl"aufige Geschwindigkeit auf den Raum der divergenzfreien Funktionen projiziert. Das klassische *Chorin-Verfahren* basiert zur Geschwindigkeitsfortpflanzung auf dem impliziten Euler-Schema. Unter Vernachl"assigung der Divergenzfreiheit wird zun"achst per

$$\frac{\tilde{v}^m - v^{m-1}}{k_m} - \nu\Delta\tilde{v}^m + \tilde{v}^m \cdot \nabla\tilde{v}^m = f^m,$$

eine Vorhersage f"ur die neue Geschwindigkeit berechnet. F"ur diese Geschwindigkeit gilt im Allgemeinen $\nabla \cdot \tilde{v}^m \neq 0$. Zur Korrektur wird $\tilde{v}^m$ im L^2-Sinne orthogonal auf den Raum H der schwach-divergenzfreien Funktionen projiziert:

$$v^m \in H := \{\phi \in L^2(\Omega)^d,\ (\phi, \nabla\xi) = 0\ \forall \xi \in C^1(\bar{\Omega})\}: \quad (v^m - \tilde{v}^m, \phi) = 0\ \forall \phi \in H.$$

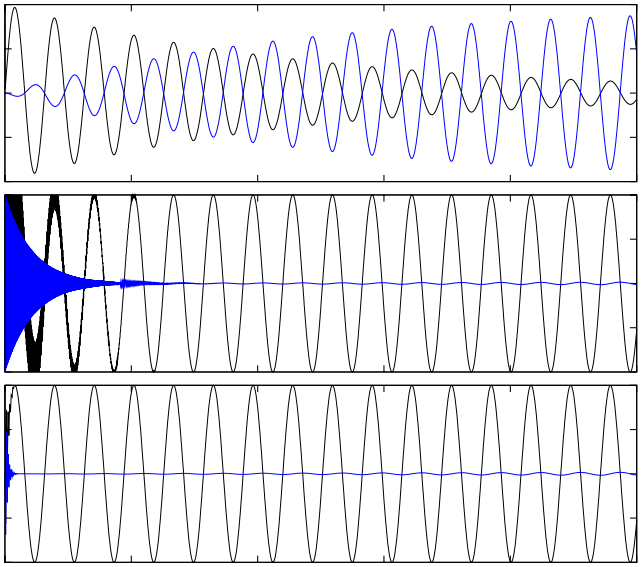


Figure 3.14: Erste L"osungskomponente (schwarz) sowie numerische Fehler (blau) bei Diskretisierung des Testproblems mit dem (von oben nach unten) impliziten Euler-Verfahren $k = 0.125/3$, Crank-Nicolson-Verfahren $k = 0.125/3$ und dem Fractional-Step-Theta-Verfahren $k = 0.125$.

Im schwachen Sinne besitzt der Raum H neben der Divergenzfreiheit auch die Randbedingung $\mathbf{n} \cdot \mathbf{v}|_{\partial\Omega} = 0$, denn

$$(\mathbf{v}, \nabla\xi)_\Omega = \langle \mathbf{n} \cdot \mathbf{v}, \xi \rangle_{\partial\Omega} - (\nabla \cdot \mathbf{v}, \xi)_\Omega = 0.$$

Diese Projektion soll berechnet werden, ohne erneut ein Sattelpunktproblem l"osen zu m"ussen. Zur Korrektur der Geschwindigkeit machen wir daher mit einem $q^m \in H^1(\Omega)$ den Ansatz:

$$\mathbf{v}^m := \tilde{\mathbf{v}}^m - k_m \nabla q^m.$$

Aus der Bedingung $\nabla \cdot \mathbf{v}^m = 0$ folgern wir die Gleichung:

$$0 = \nabla \cdot \mathbf{v}^m = \nabla \cdot \tilde{\mathbf{v}}^m - k_m \Delta q^m \quad \Leftrightarrow \quad \Delta q^m = \frac{1}{k_m} \nabla \cdot \tilde{\mathbf{v}}^m, \ \partial_n q^m = 0.$$

Die Neumann-Randbedingung wird gefordert, damit ∇q^m die Randbedingung des Raum H erf"ullt. Diese Neumann-Aufgabe hat eine L"osung, da die Kompatiblit"atsbedingung

$$\int_{\partial\Omega} \nabla \cdot \tilde{\mathbf{v}}^m dx = \int_{\partial\Omega} n \cdot \tilde{\mathbf{v}}^m ds = 0,$$

erf"ullt ist. Unter der Zusatzbedingung $(q^m, 1)_\Omega = 0$ ist die L"osung eindeutig. F"ur die korrigierte Geschwindigkeit $\mathbf{v}^m := \tilde{\mathbf{v}}^m - k_m \nabla q^m \in H^1(\Omega)^d$ gilt

$$(\mathbf{v}^m, \phi) = (\tilde{\mathbf{v}}^m, \phi) - k_m(\nabla q^m, \phi) = (\tilde{\mathbf{v}}^m, \phi) + k_m(q^m, \nabla \cdot \phi) = (\tilde{\mathbf{v}}^m, \phi) \ \forall \phi \in H.$$

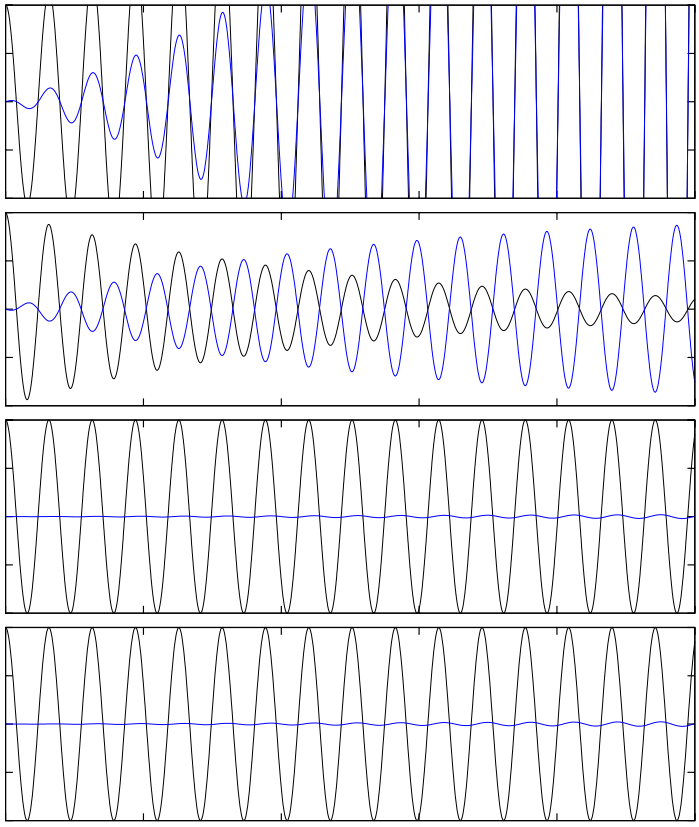


Figure 3.15: Dritte L"osungskomponente (schwarz) sowie Fehler (blau) bei Diskretisierung mit dem (von oben nach unten): expliziten Euler-Verfahren $k = 0.125/3$, impliziten Euler-Verfahren $k = 0.125/3$, Crank-Nicolson-Verfahren $k = 0.125/3$ und dem Fractional-Step-Theta-Verfahren $k = 0.125$.

Sie ist somit die gesuchte L^2-Projektion. Weiter gilt:

$$\nabla \cdot \mathbf{v}^m = \nabla \cdot \tilde{\mathbf{v}}^m - k_m \nabla \cdot \nabla q^m = \nabla \cdot \tilde{\mathbf{v}}^m - k_m \Delta q^m = 0.$$

Das Chorin-Verfahren hat allerdings den Nachteil, dass Haftrandwerte nicht exakt erf"ullt werden. Auf dem Rand gilt:

$$\mathbf{n} \cdot \mathbf{v}^m\big|_{\partial\Omega} = \mathbf{n} \cdot \tilde{\mathbf{v}}^m\big|_{\partial\Omega} - k_m \mathbf{n} \cdot \nabla q^m\big|_{\partial\Omega} = 0.$$

F"ur die Korrektur ∇q^m k"onnen jedoch keine Randwerte auf der Tangentialkomponente vorgeschrieben werden. Im Allgemeinen ist

$$\tau \cdot \mathbf{v}^m = \underbrace{\tau \cdot \tilde{\mathbf{v}}^m}_{=0}\big|_{\partial\Omega} + k_m \underbrace{\tau \cdot \nabla q^m}_{\neq 0}\big|_{\partial\Omega} \neq 0.$$

Die Hilfsl"osung q^m wird dar"uber hinaus als Druck p^m verwendet. Insbesondere in der N"ahe des Randes ist dies keine gute Approximation an den Druck. Wir fassen das Chorin-Verfahren zusammen:

Algorithmus 3.27 (Das Chorin-Verfahren)**.** Es sei $\mathbf{v}^0$ der Startwert. Auf einem Zeitgitter $0 = t_0 < t_1 < \cdots < t_M = T$, iteriere f"ur $m \geqslant 1$:

1. Geschwindigkeits-Prediktor-Schritt

$$\frac{1}{k_m}(\tilde{\mathbf{v}}^m - \mathbf{v}^{m-1}) - \nu\Delta\tilde{\mathbf{v}}^m + \tilde{\mathbf{v}}^m \cdot \nabla\tilde{\mathbf{v}}^m = \mathbf{f}^m \quad \text{in } \Omega,\ \tilde{\mathbf{v}}^m\big|_{\partial\Omega} = 0$$

2. Druck-Poisson-Gleichung:

$$\Delta q^m = \frac{1}{k_m}\nabla \cdot \tilde{\mathbf{v}}^m \quad \text{in } \Omega,\ \partial_n q^m\big|_{\partial\Omega} = 0$$

3. Korrektur

$$\mathbf{v}^m := \tilde{\mathbf{v}}^m - k_m \nabla q^m, \quad p^m := q^m$$

Dieses Verfahren verf"ugt "uber einige wesentlichen Nachteile:

- Die L"osung des Chorin-Verfahrens erf"ullt nicht die Haft-Randbedingung. Auch der "Druck" $p^m = q^m$ besitzt einen gro"sen Randfehler. Oft aber ist die L"osung gerade in der N"ahe des Randes interessant um Kr"afte zu berechnen. Hier ist das Chorin-Verfahren nicht zu gebrauchen.
- Das Chorin-Verfahren kann nicht verwendet werden um station"are Limiten zu berechnen. Selbst wenn eine L"osung $\mathbf{v}^m$ die station"are Gleichung erf"ullt, so ist die n"achste Iteration $\mathbf{v}^{m+1}$ eine St"orung derselben.
- Das Verfahren ist nur von niedriger Zeit-Ordnung. Obwohl das Verfahren auf dem impliziten Euler-Verfahren basiert, betr"agt die Ordnung lediglich $O(\sqrt{k})$. Abhilfe schaffen Modifikationen, welche nicht auf dem impliziten Euler, sondern z.B. auf dem Crank-Nicolson Verfahren aufbauen wie z.B. das *Van Kan-Verfahren*. Dieses Verfahren kann auch verwendet werden, um station"are Limiten zu berechnen.

Analysen zur Chorin-Methode sowie Details zu weiteren Projektionsmethoden finden sich in der Literatur [**?**].

4 Adaptive Finite Elemente f”ur die Navier-Stokes Gleichungen

Insbesondere in drei Raumdimensionen ist die Verwendung von lokal verfeinerten Gittern zur Diskretisierung der Navier-Stokes Gleichungen unerl”asslich. Ziel einer *adaptiven Gittersteuerung* ist das automatische Verfeinern der Gitter. Zur Verfeinerung werden *Fehlerindikatoren* η_K hergeleitet, welche f”ur jedes Element $K \in \Omega_h$ des Gitters den lokalen Fehleranteil beschreiben. Diese *Fehlerindikatoren* entstehen durch *Lokalisierung* eines *Fehlersch”atzers* η:

$$\eta \sim J(u) - J(u_h).$$

Im Gegensatz zu einfachen Residuuen-Fehlersch”atzern, welche den Fehler ”uberlicherweise in der *Energie-Norm*, also z.B.

$$\nu\|\nabla(\mathbf{v} - \mathbf{v}_h)\| + \|p - p_h\|,$$

sch”atzen, wollen wir allgemeine *Fehlerfunktionale* $J(\cdot) : V \to \mathbb{R}$ betrachten. Im Kontext der numerischen Str”omungsmechanik kann so ein Funktional z.B. die auf ein Hindernis wirkende Kraft sein

$$J(u_h) = \int_{\Gamma_o} \mathbf{n} \cdot \boldsymbol{\sigma} \cdot \mathbf{e}\, ds = \int_{\Gamma_o} (\nu \partial_n \mathbf{v}_h - \mathbf{n} \cdot \mathbf{v}_h) \cdot \mathbf{e}\, ds,$$

wobei Γ_o der Rand des Hindernis, n der Normalvektor an Γ_o und e die Richtung ist, in der die Kraft gemessen wird. Ein *adaptiver Algorithmus* l”auft wie folgt:

Algorithmus 4.1 (Adaptiver Finite-Elemente Algorithmus)**.**

Sei mit Ω_h eine Anfangstriangulierung von Ω gegeben. Iteriere:

1. L”ose die Differentialgleichung $u_h \in V_h$ auf Ω_h
$$A(u_h)(\Phi_h) = F(\Phi_h) \quad \forall \Phi_h \in V_h$$
2. Sch”atze den Fehler
$$\eta(u_h) \sim J(u) - J(u_h)$$
3. Falls $\eta < \epsilon$ Abbruch
4. Erstelle lokale Fehlerindikatoren
$$\eta(u_h) \sim \sum_{K \in \Omega_h} \eta_K(u_h, K)$$
5. Verfeinere das Gitter $\Omega_h \to \Omega_{h'}$ wo η_K gro”s, weiter bei 1 mit $\Omega_{h'}$

4.1 Die DWR-Methode

Wir beschreiben die DWR-Methode (*dual weighted residual*-method) zur Fehlersch"atzung zun"achst in einem abstrakten Rahmen. Es sei durch $u \in V$:

$$A(u)(\phi) = F(\phi) \quad \forall \phi \in V, \tag{4.1}$$

eine partielle Differentialgleichung gegeben. Dabei sei $A(\cdot)(\cdot) : V \times V \to \mathbb{R}$ eine Semilinearform, linear im zweiten Argument, sowie $F(\cdot) : V \to \mathbb{R}$ ein lineares Funktional. Auf einer Triangulierung Ω_h des Gebiets Ω wird die diskrete L"osung $u_h \in V_h$ gesucht als L"osung von

$$A(u_h)(\phi_h) = F(\phi_h) \quad \forall \phi_h \in V_h. \tag{4.2}$$

Wir wollen den Fehler in einem (nicht notwendigerweise linearem) Funktional $J(\cdot) : V \to \mathbb{R}$ bestimmen. Wir formulieren das Problem als ein *triviales Optimierungsproblem mit Nebenbedingung*:

$$J(u) \to \min!, \quad \text{wobei} \quad A(u)(\phi) = F(\phi) \quad \forall \phi \in V.$$

Wir gehen davon aus, dass die Probleme (4.1) und (4.2) genau eine L"osung haben. Daher wird die Nebenbedingung nur von der L"oung u selbst erf"ullt und die Minimierung erstreckt sich "uber die triviale Menge $\{u\}$.

Das zu diesem Minimierungsproblem geh"orige Lagrange-Funktional ist gegeben durch

$$L(u, z) := J(u) + F(z) - A(u)(z). \tag{4.3}$$

Wir nennen $z \in V$ die duale L"osung. Entsprechend definieren wir das diskrete Lagrange-Funktional als

$$L(u_h, z_h) := J(u_h) + F(z_h) - A(u_h)(z_h). \tag{4.4}$$

Die L"osung $u \in V$ und $z \in V$ des Minimierungsproblems ist durch den station"aren Punkt des Lagrange-Funktionals gegeben ($F(\cdot)$ ist linear und $A(\cdot)(\cdot)$ ist linear im zweiten Argument):

$$L'(u, z)(\delta u, \delta z) := J'(u)(\delta u) + F(\delta z) - A'(u)(\delta u, z) - A(u)(\delta z) = 0 \quad \forall \delta u, \delta z \in V,$$

wobei die Richtungsableitung der Form definiert ist wie in (3.24), bei der Verwendung f"ur das Newton-Verfahren. Es ergeben sich f"ur $u, z \in V$ und $u_h, z_h \in V_h$ die Gleichungen des *primalen* und *dualen* Problems:

$$\begin{array}{rlrl} A(u)(\phi) = F(\phi) & \forall \phi \in V & A'(u)(\phi, z) = J'(u)(\phi) & \forall \phi \in V \\ A(u_h)(\phi_h) = F(\phi_h) & \forall \phi_h \in V_h & A'(u_h)(\phi_h, z_h) = J'(u_h)(\phi_h) & \forall \phi_h \in V_h. \end{array} \tag{4.5}$$

Das primale Problem stimmt gerade mit der Differentialgleichung (4.1), bzw. (4.2) "uberein. Sind $u \in V$ und $u_h \in V_h$ L"osungen des primalen Problems, so gilt f"ur beliebige $\phi \in V$ und $\phi_h \in V_h$ die Darstellung:

$$J(u) - J(u_h) = L(u, \phi) - L(u_h, \phi_h).$$

Diese Identit"at soll genutzt werden, um einen Fehlersch"atzer herzuleiten. Zur Vereinfachung fassen wir zusammen

$$x := (u, z) \in X := V \times V, \quad x_h = (u_h, z_h) \in X_h := V_h \times V_h,$$

und k"onnen also f"ur den Fehler ausdr"ucken als

$$J(u) - J(u_h) = L(x) - L(x_h).$$

Wir erhalten elementar die folgende Fehlerdarstellung:

Satz 4.2 (Fehleridentit"at)**.** F"ur die Galerkin-L"osungen $x = (u, z) \in X$ und $x_h = (u_h, z_h) \in X_h$ von (4.5) gilt die *a posteriori Fehleridentit"at*

$$J(u) - J(u_h) = L(x) - L(x_h) = \frac{1}{2}L'(x)(x - i_h x) + R. \tag{4.6}$$

Das Restglied R ist von dritter Ordnung im Fehler $e := x - x_h$

$$R := \frac{1}{2}\int_0^1 L'''(x_h + se)(e, e, e)s(s-1)ds, \tag{4.7}$$

und ist Null $R = 0$, falls das Lagrange-Funktional $L(\cdot)$ quadratisch ist.

Beweis: Mit dem Hauptsatz der Integralrechnung erhalten wir eine Integraldarstellung des Fehlers $e := x - x_h$:

$$L(x) - L(x_h) = \int_0^1 L'(x_h + se)(e)\,ds$$

Dieses Integral approximieren wir mit der Trapez-Regel. Mit dem Restglied (4.7) (dies ist gerade das Restglied der Trapez-Regel) gilt

$$L(x) - L(x_h) = \int_0^1 L'(x_h + se)(e)\,ds = \frac{1}{2}L'(x_h)(e) + \frac{1}{2}L'(x)(e) + R.$$

Wir nutzen die Ableitung von (4.4) und erhalten mit (4.5):

$$\begin{aligned} L'(x_h)(x - x_h) &= \underbrace{F(z - z_h) - A(u_h)(z - z_h)}_{primal} + \underbrace{J'(u_h)(u - u_h) - A'(u_h)(u - u_h, z_h)}_{dual} \\ &= F(z - i_h z) - A(u_h)(z - i_h z) + J'(u_h)(u - i_h u) - A'(u_h)(u - i_h u, z_h) \\ &= L'(x_h)(x - i_h x). \end{aligned}$$

Im zweiten Term kann die Orthogonalit"at der kontinuierlichen L"osung $u \in V$ und $z \in V$ genutzt werden. Es gilt wegen $x - x_h \in X$:

$$L'(x)(x - x_h) = \underbrace{F(z - z_h) - A(u, z - z_h)}_{=0} + \underbrace{J'(u)(u - u_h) - A'(u)(u - u_h, z)}_{=0} = 0$$

Damit ist dieser einfache Satz bewiesen. □

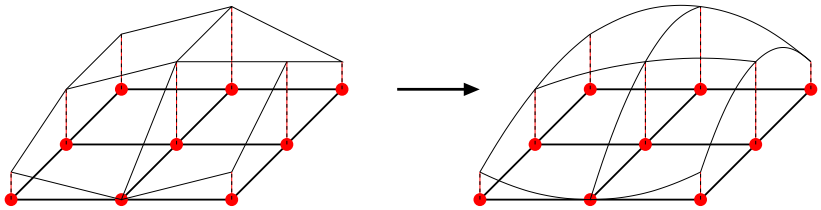


Figure 4.1: Interpolation $i_h^* : X_h \to X_h^*$.

Dieses Ergebnis erlaubt es, den Funktionalfehler $J(u) - J(u_h)$ "uber die Residuuen des primalen und dualen Problems darzustellen (4.6). Im folgenden Abschnitt konkretisieren wir die DWR-Methode f"ur die Navier-Stokes Gleichungen.

Um die Fehleridentit"at als Fehlersch"atzer auswerten zu k"onnen

$$\eta_h(u_h, z_h) := F(z - i_h z) - A(u_h)(z - i_h z) + J'(u_h)(u - i_h u) - A'(u_h)(u - i_h u, z_h),$$

ist eine Approximation f"ur die Interpolationsfehler $u - i_h u$ und $z - i_h z$ notwendig. Denn die kontinuierlichen L"osungen $u \in V$ und $z \in V$ sind nicht bekannt. Im folgenden stellen wir einige Zug"ange zur Approximation vor.

1. Die diskreten L"osungen $x_h = (u_h, z_h) \in X_h$ k"onnten zus"atzlich mit h"oherer Genauigkeit $x_h^* \in X_h^*$ berechnet werden. Dabei kann X_h^* entweder ein Finite-Elemente Raum auf feinerem Gitter, etwa $h' = h/2$ oder ein Finite-Elemente Raum mit h"oherem Ansatzgrad sein (z.B. Q^2 statt Q^1-Elemente).

   $$\eta_h^*(u_h, z_h, u_h^*, z_h^*) := F(z_h^* - i_h z_h^*) - A(u_h)(z_h^* - i_h z_h^*) + J'(u_h)(u_h^* - i_h u_h^*) - A'(u_h)(u_h^* - i_h u_h^*, z_h),$$

   Beide Vorgehen sind jedoch zu kostspielig und w"urden bedeuten, dass zum Sch"atzen des Fehler mehr Aufwand betrieben wird, als zum L"osen des eigentlichen Problems. Dieser Fehlersch"atzer liefert sehr gute Resultate, f"ur den Effektivit"atsindex gilt

   $$I_{\text{eff}} := \frac{\eta_h^*(u_h, z_h, u_h^*, z_h^*)}{|J(u) - J(u_h)|} \to 1 \quad (h \to 0).$$

2. Eine weitere Alternative unmittelbar die Fehleridentit"at zu verwenden besteht darin, den Interpolationsfehler durch *Superapproximation* zu n"ahern. Statt mit $(u_h^*, z_h^*) \in X_h^*$ eine L"osung von h"oherem Polynomgrad zu berechnen, konstruieren wir einen diskreten Interpolationsoperator $i_h^* : X_h \to X_h^*$. Dies geschieht lokal durch "Anderung der Finite Elemente Basis. Hat $u_h \in V_h$ die Darstellung

   $$u_h = \sum_{i=1}^{N} \mathbf{u}_i \phi_i,$$

   mit der Basis $\{\phi_i, i = 1, \ldots, N\}$ des X_h, so ist die Interpolation i_h^* gegeben durch

   $$i_h^* u_h = \sum_{i=1}^{N} \mathbf{u}_i \phi_i^*,$$

wobei $\{\phi_i^*, i = 1, \ldots, N\}$ die Basis des X_h^* ist. F"ur X_h^* w"ahlen wir den Raum mit doppeltem Polynomgrad auf dem Gitter $h' = 2h$. F"ur diesen Raum gilt $\dim(X_h) = \dim(X_h^*)$ und die Knotenfunktionale liegen am gleichen Ort. In Abbildung 4.1 ist die diskrete Interpolation vom Raum Q^1 auf Ω_h in den Raum Q^2 auf Ω_{2h} an einem Beispiel dargestellt. Diese Approximation kann ohne erneutes Rechnen erfolgen. F"ur die Auswertung des Fehlersch"atzers m"ussen die Residuen lediglich mit h"oherer Quadraturforlem ausgewertet werden:

$$\eta_h^*(u_h, z_h) := F(i_h^* z_h - z_h) - A(u_h)(i_h^* z_h - z_h) + J'(u_h)(i_h^* u_h - u_h) - A'(u_h)(i_h^* u_h - u_h, z_h). \tag{4.8}$$

Dieses Vorgehen f"uhrt in der Regel zu sehr guten Fehlersch"atzern η_h^*. Bei hinreichender Regularit"at gilt

$$I_{\text{eff}} := \frac{\eta_h(u_h, z_h)}{|J(u) - J(u_h)|} \to 1 \quad (h \to 0).$$

Eine theoretische Analyse ist hingegen schwierig, das Vorgehen macht sich Superapproximationseffekte der Finite-Elemente L"osung u_h in den Knotenpunkten zu Nutze.

3. Die Residuen k"onnen mit partieller Integration und der Cauchy-Schwarz Ungleichung weiter abgesch"atzt werden. Diese Darstellung h"angt von der Gleichung ab. Der Fehlersch"atzer hat dann die Form

$$\eta_h := \sum_{K \in \Omega_K} \{\rho_K(u_h) \cdot \omega_K(z_h) + \rho_K^*(u_h, z_h)\omega_K^*(u_h)\},$$

mit *primalen Residuen* $\rho_K(u_h)$ und *Gewichten* $\omega_K(z_h)$ sowie *dualen Residuen* $\rho_K^*(u_h, z_h)$ und *dualen Gewichten* $\omega_K^*(u_h)$. Die Gewichte sind Interpolationsfehler auf der Zelle K, welche dann mittels Interpolationsabsch"atzungen approximiert werden k"onnen:

$$\|u - i_h u\|_K \leqslant c_i h_K^2 \|\nabla^2 u\|_K \sim h_K^2 \|\nabla_h^2 u_h\|_K.$$

Wir werden dieses Vorgehen f"ur die Navier-Stokes Gleichungen konkretisieren. Wegen der groben Absch"atzung des Fehlers mit der Cauchy-Schwarz Ungleichung liefert dieser einfache Fehlersch"atzer "ublicherweise eine starke "Ubersch"atzung des Fehlers und es gilt

$$I_{\text{eff}} := \frac{\eta_h(u_h, z_h)}{|J(u) - J(u_h)|} \gg 1.$$

Desweiteren ist die Interpolationskonstante nicht bekannt und es muss $c_i \sim 1$ verwendet werden. Die lokalen Gr"o"sen

$$\eta_K(u_h, z_h) := \rho_K(u_h)\omega_K(z_h) + \rho_K^*(u_h, z_h)\omega_K^*(u_h)$$

lassen sich jedoch gut als Verfeinerungsindikatoren zur Gittersteuerung verwenden.

4.2 Fehlersch"atzung bei den Navier-Stokes Gleichungen

Die Navier-Stokes Gleichungen sind mit $U = (v, p) \in H_0^1(\Omega)^d \times L_0^2(\Omega)$ durch die Semilinearform

$$A(U)(\Phi) = \nu(\nabla v, \nabla\phi) + (v\cdot\nabla v, \phi) - (p, \nabla\cdot\phi) + (\nabla\cdot v, \xi) = (f, \phi), \quad \forall\Phi := (\phi, \xi) \in H_0^1(\Omega)^d \times L_0^2(\Omega)$$

gegeben. Als *Zielfunktional* betrachten wir den Widerstands- (*Drag*) c_D sowie die Auftriebs- (*Lift*) Koeffizienten c_L an einem umstr"omten Hindernis mit Rand Γ_o:

$$J_D(u) := c_D := \frac{2}{\rho V^2 L}\int_{\Gamma_o} (\partial_n v - p\cdot n)\cdot e_1 \, ds, \quad J_L(u) := c_L := \frac{2}{\rho V^2 L}\int_{\Gamma_o} (\partial_n v - p\cdot n)\cdot e_2 \, ds.$$

Hier ist n der nach au"sen gerichte Normalvektor an Γ_o und $e_1 = (1,0)^T$, sowie $e_2 = (0,1)^T$ sind die Einheitsvektoren in Hauptstr"omungsrichtung bzw. orthogonal zur Hauptstr"omungsrichtung. Das zugeh"orige Lagrange-Funktional mit $Z = (w, q) \in H_0^1(\Omega)^d \times L_0^2(\Omega)$ gegeben durch

$$L(U, Z) = J(U) + F(Z) - A(U)(Z).$$

Primale und duale L"osung sind durch die Nullstellen der Ableitungen bestimmt. Die Primale L"osung (Ableitung nach Z) ist durch die Navier-Stokes Gleichungen selbst gegeben:

$$A(U)(\Phi) = F(\Phi) \quad \forall\Phi \in H_0^1(\Omega)^d \times L_0^2(\Omega).$$

F"ur die duale Gleichung muss die Ableitung der Semilinearform bestimmt werden:

$$A'(U)(\delta U, Z) = \nu(\nabla\delta v, \nabla w) + (v\cdot\nabla\delta v, w) + (\delta v\cdot\nabla v, w) - (\delta p, \nabla\cdot w) + (\nabla\cdot\delta u, q).$$

Das Funktional hat die Ableitung (am Beispiel des Drag)

$$J_D'(u)(\Phi) = \int_{\Gamma_o} (\partial_n\phi - n\cdot p)\cdot e_1 \, ds = \underbrace{\int_{\Gamma_o} \partial_n\phi^1 \, ds}_{=:J_v'(\phi)} \underbrace{- \int_{\Gamma_o} n_x \xi \, ds}_{=:J_p'(\xi)}.$$

Die duale L"osung $Z = (w, q)$ ist bestimmt durch

$$\begin{aligned} \nu(\nabla w, \nabla\phi) + (w, v\cdot\nabla\phi) + (w, \phi\cdot\nabla v) + (q, \nabla\cdot\phi) &= J_v'(\phi) \quad \forall\phi \in H_0^1(\Omega)^d \\ -(\nabla\cdot w, \xi) &= J_p'(\xi) \quad \forall\xi \in L_0^2(\Omega). \end{aligned} \tag{4.9}$$

Der *duale Konvektionsterm* kann (bei Dirichlet-Randwerten) durch partielle Integration transformiert werden zu:

$$(w, v\cdot\nabla\phi) + (w, \phi\cdot\nabla v) = -(v\cdot\nabla w, \phi) + \begin{pmatrix} v_x^2 w^2 - v_y^2 w^1, \phi^1 \\ v_y^1 w^1 - v_x^1 w^2, \phi^2 \end{pmatrix}.$$

Er spaltet sich also in einen Transport-Term (der erste) in Richtung $-v$, also entgegen der Str"omungsrichtung und in einen Reaktionsterm nullter Ordnung auf. Auch der Druck erscheint mit umgekehrtem Vorzeichen. Vernachl"assigen wir den Konvektionsterm, beschr"anken

uns also auf die Stokes-Gleichungen, so kann die klassische Formulierung der dualen Gleichung angegeben werden:

$$\begin{aligned} -\nu\Delta w - \nabla q &= j_\nu \\ -\nabla \cdot w &= j_p, \end{aligned}$$

wobei $j_\nu, j_p \in H^{-1}(\Omega)$ die *klassischen Darstellungen* der Funktionale sind, gegeben durch

$$(j_\nu, \phi) = J'_\nu(\phi)\ \forall \phi \in H_0^1(\Omega)^d, \quad (j_p, \xi) = J'_p(\xi)\ \forall \phi \in L_0^2(\Omega).$$

Die duale L"osung $z = (w, q)$ kann als *Sensitivit"at* der L"osung bez"uglich des Funktionals betrachtet werden. Sie gibt an, welche Bereiche der L"osung wesentlich f"ur die Funktionalauswertung sind.

Der Fehlersch"atzer l"asst sich nun angeben als

$$J(u,p) - J(u_h,p_h) \approx \eta_h := \frac{1}{2}\rho(\nu_h,p_h)(w - i_h w, q - i_h q) + \frac{1}{2}\rho^*(w_h,q_h)(\nu - i_h\nu, p - i_h p),$$

mit den Residuen des primalen und dualen Problems. Wir konkretisieren dies zur Vereinfachung f"ur die Stokes-Gleichungen

$$\begin{aligned} \rho(\nu_h,p_h)(\delta w,\delta q) &:= F(\delta w) - \nu(\nabla u_h,\nabla\delta w) + (p_h,\nabla\cdot\delta w) + (\nabla\cdot\nu_h,\delta q), \\ \rho^*(u_h,p_h)(w_h,q_h)(\delta\nu,\delta p) &:= J'_\nu(\delta\nu) - \nu(\nabla w_h,\nabla\delta\nu) + (q_h,\nabla\cdot\delta\nu) \\ &\quad + J'_p(\delta p) + (\nabla\cdot w_h,\delta p). \end{aligned}$$

Um den Fehlersch"atzer als *a posteriori* Fehlersch"atzer auswerten zu k"onnen m"ussen die Interpolationsfehler $\delta\nu := \nu - i_h\nu$, $\delta p := p - i_h p$, sowie $\delta w := w - i_h w$ und $\delta q := q - i_h q$ wie oben beschrieben approximiert werden. F"ur einen effizienten Fehlersch"atzer sollte die diskrete Interpolation $i_h^* : X_h \to X_h^*$ verwendet werden. Dies erfordert jedoch eine gewisse Gitter-Struktur, welche nicht in jeder Finite-Elemente Software bereitgestellt wird. Wir beschreiben daher die dritte Variante. F"ur die primalen Residuuen gilt bei partieller Integration:

$$\rho(\nu_h,p_h)(\delta w,\delta q) = \sum_{K\in\Omega_h}\Big\{(f + \nu\Delta\nu_h - \nabla p_h,\delta w)_K - \langle\nu\partial_n\nu_h,\delta w\rangle_{\delta K} + (\nabla\cdot\nu_h,\delta q)_K\Big\}.$$

Wir sch"atzen die Terme mit der Cauchy-Schwarz Ungleichung ab und fassen das Randintegral zu Spr"ungen zusammen:

$$e = K\cap K' : \quad [\partial_n\nu_h]_e := (\partial_n\nu_h\big|_K - \partial_n\nu_h\big|_{K'})\Big|_e.$$

Wir erhalten:

$$|\rho(\nu_h,p_h)(\delta w,\delta q)| \leqslant \sum_{K\in\Omega_h}\Big\{\|f+\nu\Delta\nu_h-\nabla p_h\|_K\,\|\delta w\|_K + \frac{1}{2}\|\nu[\partial_n\nu_h]\|_{\partial K}\,\|\delta w\|_{\partial K} + \|\nabla\cdot\nu_h\|_K\,\|\delta q\|_K\Big\}.$$

Angenommen, wir verwenden das Taylor-Hood Element mit st"uckweise quadratischen Geschwindigkeiten v_h und st"uckweise linearem Druck p_h, so gilt weiter mit $\delta w := w - i_h w$ und $\delta q := q - i_h q$

$$\|w - i_h w\| + h_K^{\frac{1}{2}} \|w - i_h w\|_{\partial K} \leqslant c_i h_K^3 \|\nabla^3 w\|, \quad \|q - i_h q\| \leqslant c_i h_K^2 \|\nabla^2 q\|_K.$$

Damit k"onnen die primalen Residuuen abgesch"atzt werden zu

$$|\rho(v_h, p_h)(\delta w, \delta q)| \leqslant c_i \sum_{K \in \Omega_h} \left\{ h_K^3 \left(\|f + \nu \Delta v_h - \nabla p_h\|_K + \frac{1}{2} h_K^{-\frac{1}{2}} \|\nu[\partial_n v_h]\|_{\partial K} \right) \|\nabla^3 w\|_K \right.$$
$$\left. + h_K^2 \|\nabla \cdot v_h\|_K \, \|\nabla^2 q\|_K \right\}.$$

F"ur die dualen Residuuen gilt entsprechend:

$$|\rho^*(w_j, q_h, v_h, p_h)(\delta v, \delta p)| \leqslant c_i \sum_{K \in \Omega_h} \left\{ h_K^3 \left(\|j_v + \nu \Delta w_h + \nabla q_h\|_K + \frac{1}{2} h_K^{-\frac{1}{2}} \|\nu[\partial_n w_h]\|_{\partial K} \right) \|\nabla^3 v\|_K \right.$$
$$\left. + h_K^2 \|j_p + \nabla \cdot w_h\|_K \, \|\nabla^2 p\|_K \right\}.$$

Die h"oheren Ableitungen der L"osungen m"ussen durch geeignete lokale Differenzenquotienten abgesch"atzt werden. Diese Darstellung des Fehlersch"atzers eignet sich gut um lokal verfeinerte Gitter zu erstellen, es liegt im Allgemeinen jedoch eine Starke "Ubersch"atzung des Fehlers vor.

4.3 Strategien zur Gitterverfeinerung

Basis einer adaptiven Gitteradaption sind *lokale Fehlerindikatoren*. Wir ben"otigen eine lokale Darstellung des Fehlers der Art

$$|\eta_h| \sim \sum_{K \in \Omega_h} \eta_K,$$

mit Indikatoren η_K auf jedem Element der Triangulierung. Bei der Lokalisierung des Fehlersch"atzers ist darauf zu achten, dass die Fehlerindikatoren die richtige Ordnung wiedergeben. Werden die Integrale bei der Berechnung des Fehlsch"atzers (4.8) zur Lokalisierung einfach auf die einzelnen Zellen der Diskretisierung eingeschr"ankt, so liegt lokal eine zu geringe Approximationsordnung vor und es wird zuviel verfeinert.

Alle Verfeinerungsstrategien suchen Elemente mit gro"sem Indikatorwert η_K zur Verfeinerung aus. Wir beschreiben hier einige einfache Methoden

1. *Fixed Number* oder *Fixed Fraction* Stratiegie. Die Indikatorwerete werden absteigend sortiert

$$\eta_{K_1} \geqslant \eta_{K_2} \geqslant \ldots \eta_{K_N}.$$

Bei der *Fixed Number* Strategie werden die ersten $r_\alpha := \alpha N$ Elemente mit $\alpha \in (0,1)$ verfeinert:

$$\{K_i,\ i = 1, \dots, r_\alpha\}.$$

Bei der *Fixed Fraction* Stratiegie werden die Elemente mit gr"o"sten Fehlerindikatorwerten verfeinert, welche insgesamt einen bestimmten Anteil des Gesamtfehlers ausmachen:

$$\{K_i,\ i = 1, \dots, r_\alpha,\ \sum_{i=1}^{r_\alpha} \eta_{K_i} \leqslant \eta \sum_{i=1}^{N} \alpha_{K_i}\}$$

Beide Strategien haben den Nachteil, dass der Parameter α willk"urlich gew"ahlt werden muss.

2. *Fehler"aquilibrierungs-Strategie*. Wir nehmen an, dass bei einem optimal gesteuertem Gitter die Fehler "aquilibriert sind, dass also

$$\eta_K \approx \eta_{K'}$$

gilt f"ur alle Elemente $K, K' \in \Omega_h$. Um dieses Ziel zu erreichen verfeinern wir s"amtliche Elemente mit einem Fehlerindikatorwert gr"o"ser dem Durschschnitt:

$$\{K, \eta_K \geqslant \bar{\eta}\}, \quad \bar{\eta} := \frac{1}{N} \sum_{K \in \Omega} \eta_K.$$

Diese einfache Strategie kommt ohne weitere Parameter aus und liefert im Allgemeinen optimale Resultate.

4.4 Numerisches Beispiel: Widerstandsberechnung an einem Hindernis

Wir betrachten die Str"omung um ein Hindernis mit Rand Γ_o und wollen die auftretenden Kr"afte bestimmen. Dabei betrachten wir die Widerstandskraft F_{drag} sowie die Auftriebskraft F_{lift}, gegeben durch

$$F_{drag} = \int_{\Gamma_o} n \cdot \sigma_f \cdot e_1 \, ds, \quad F_{lift} = \int_{\Gamma_o} n \cdot \sigma_f \cdot e_2 \, ds,$$

mit den Einheitsvektoren $e_1 = (1,0)^T$ sowie $e_2 = (0,1)^2$. Als Kenngr"o"sen werden in der Str"omungsmechanik die dimensionslosen Gr"o"sen *Drag-* sowie *Lift-Koeffizient* c_L und c_D verwendet:

$$c_D = \frac{2F_{drag}}{\rho V^2 L}, \quad c_L = \frac{2F_{lift}}{\rho V^2 L},$$

wobei V die durchschnittliche Geschwindigkeit und L die Gr"o"se des Hindernis ist. Als Fehlerfunktionale verwenden wir also

$$J_{drag} = \frac{2}{\rho V^2 L} \int_{\Gamma_o} \left(\nu \partial_n v_h^1 - p_h n_x \right) ds, \quad J_{lift} = \frac{2}{\rho V^2 L} \int_{\Gamma_o} \left(\nu \partial_n v_h^2 - p_h n_y \right) ds.$$

Wir betrachten die Str"omung um das Forschungs-Uboot wie in Kapitel **??** und w"ahlen

$$V = 0.5,\ \nu = 0.025,\ D = 5 \quad \Rightarrow \quad Re = 100.$$

Zur Diskretisierung verwenden wir $Q^1 - Q^1$ Finite Elemente, stabilisiert mit lokalen Projektionen. F"ur Drag- und Lift-Koeffizienten ermitteln wir durch Rechnung auf sehr feinen Gittern die Referenzwerte

$$c_D = 1.24603 \pm 10^{-5}, \quad c_L = -0.2472 \pm 10^{-4}.$$

In Tabelle 4.1 fassen wir zun"achst die Werte zusammen, welche bei globaler Gitterverfeinerung ermittelt werden k"onnen.

N	$J_{drag}(u_h)$	$J_{drag}(u - u_h)$	$J_{lift}(u_h)$	$J_{lift}(u - u_h)$
842	1.5835	0.33746	-1.1435	-0.89635
3 228	1.2850	0.03900	-0.3118	-0.06457
12 632	1.2426	-0.00341	-0.2122	0.03498
49 968	1.2440	-0.00206	-0.2398	0.00731
198 752	1.2455	-0.00056	-0.2457	0.00151
792 768	1.2459	-0.00014	-0.2468	0.00034

Table 4.1: Drag- und Lift-Koeffizient bei globaler Gitterverfeinerung.

In Tabelle 4.2 verwenden wir einen Energiefehlersch"atzer in der Norm

$$\nu^{\frac{1}{2}} \|\nabla(\nu - \nu_h)\| + \|p - p_h\|$$

zur Fehlersch"atzung und Gitterverfeinerung. Dieser Fehlersch"atzer liefert nat"urlich die gleichen Gitter f"ur Drag und Lift, da das Funktional nicht eingeht.

N	$J_{drag}(u_h)$	$J_{drag}(u - u_h)$	$J_{lift}(u_h)$	$J_{lift}(u - u_h)$
842	1.5835	0.33746	-1.1435	-0.89635
1 560	1.2869	0.04086	-0.3117	-0.06450
3 208	1.2436	-0.00246	-0.2123	0.03489
7 250	1.2440	-0.00200	-0.2407	0.00648
16 426	1.2453	-0.00073	-0.2454	0.00176
39 848	1.2459	-0.00012	-0.2466	0.00054

Table 4.2: Drag- und Lift-Koeffizient bei Verwendung des Energiefehlersch"atzers.

Schlie"slich wenden wir die DWR-Methode auf dieses Problem an. Um den Fehlersch"atzer

$$\eta_h^*(u_h, z_h),$$

t

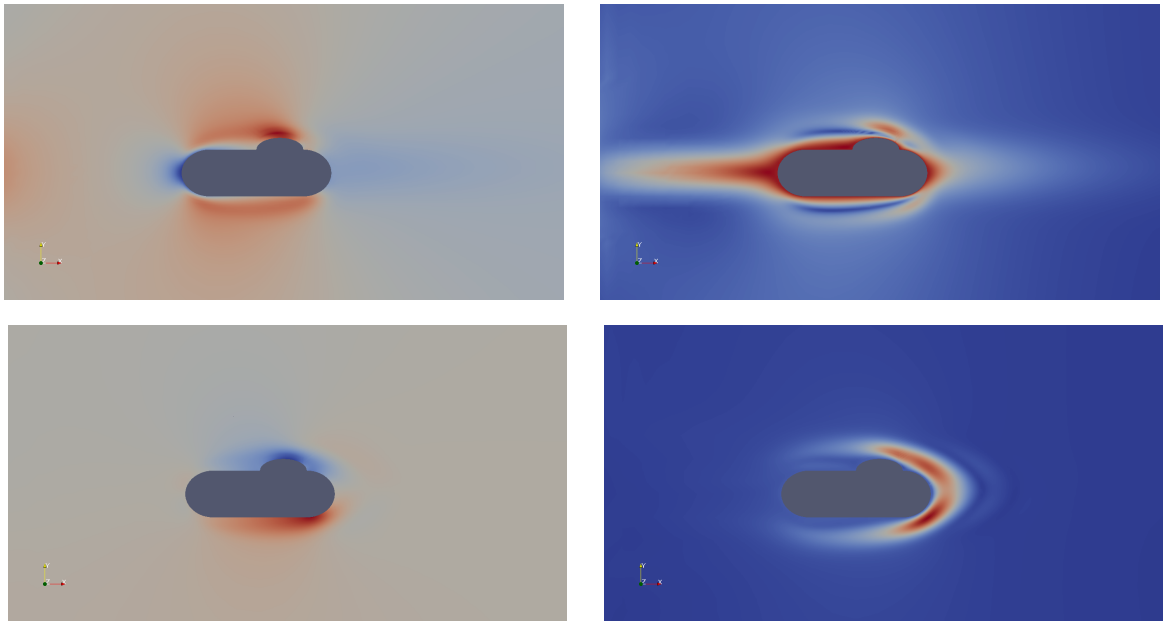


Figure 4.2: Duale L"osungen zur Widerstandsberechnung. Obere Zeile: dualer Druck (links) und duale Geschwindigkeit zum Drag, untere Zeile: dualer Druck zum Lift und duale Geschwindigkeit zum Lift.

auswerten zu k"onnen muss neben dem primalen Problem auch das duale Problem gel"ost werden. In Tabelle 4.3 fassen wir die ermittelten Werte zusammen und in Abbildung 4.2 zeigen wir duale L"osungen zu beiden Funktionalen.

N_{drag}	$J_{drag}(u_h)$	$J_{drag}(u-u_h)$	N_{lift}	$J_{lift}(u_h)$	$J_{lift}(u-u_h)$
842	1.5835	0.337459	842	-1.1435	-0.89635
1698	1.2865	0.040455	1 672	-0.3114	-0.06419
3222	1.2434	-0.002637	2 884	-0.2113	0.03584
6606	1.2443	-0.001748	5 540	-0.2405	0.00663
14808	1.2456	-0.000437	13 622	-0.2462	0.00102
36132	1.2460	-0.000022	38 136	-0.2470	0.00014

Table 4.3: Drag- und Lift-Koeffizient bei Verwendung des DWR-Sch"atzers.

Abschlie"send fassen wir alle Ergebnisse in Abbildung 4.3 zusammen. In der folgenden Tabelle vergleichen wir die ben"otigte Diskretisierungsgenauigkeit um einen Fehler von 0.1% im Drag und 0.1% im Lift zu erreichen :

	Global	Energie	DWR
Drag	192 752	16 426	6 606
Lift	792 768	39 848	13 622

Table 4.4: Ben"otigte Gittergr"o"se um einen Fehler kleiner als 0.1% im Drag, beziehungsweise 0.1% im Lift zu erreichen.

t

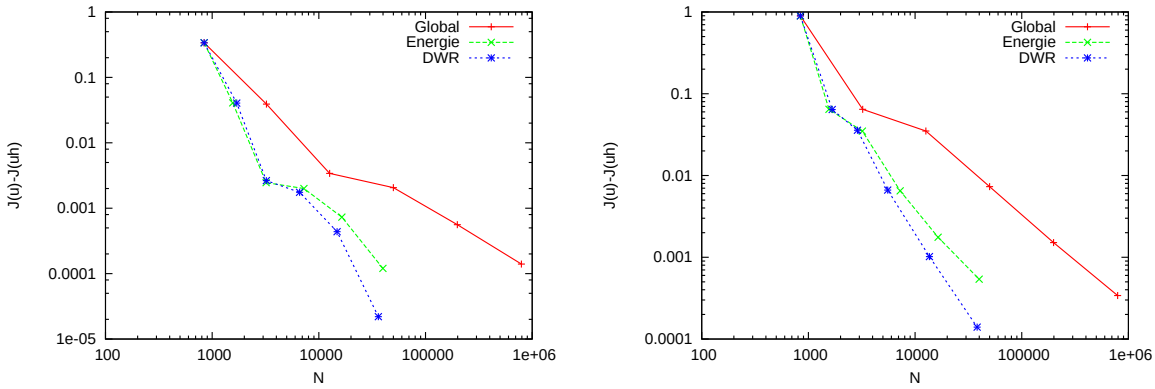


Figure 4.3: Konvergenzvergleich f"ur den Drag (links) und den Lift (rechts) bei Verwendung von globaler Gitterverfeinerung, Verfeinerung mit dem Energie-Sch"atzer und bei der DWR-Methods.

5 Solution methods

We consider inf-sub stable finite element discretizations of the Stokes and the Navier-Stokes equations. For this, let us recall the notation in Chapter 3: Let the discrete finite element spaces $V_h \times L_h$ be given as:

$$V_h := \text{span}\Big\{\phi_h^i,\ i = 1, \dots, N_h^v\Big\}, \quad L_h := \text{span}\Big\{\xi_h^i,\ i = 1, \dots, N_h^p\Big\}. \tag{5.1}$$

Then, every function $\mathbf{v}_h \in V_h$ and $p_h \in L_h$ is uniquely given as:

$$\mathbf{v}_h = \sum_{i=1}^{N_h^v} \mathbf{v}_i \phi_h^i, \quad p_h = \sum_{i=1}^{N_h^p} \mathbf{p}_i \xi_h^i, \quad \mathbf{v} \in \mathbb{R}^{N_h^v}\ \mathbf{p} \in \mathbb{R}^{N_h^p}. \tag{5.2}$$

Let us recall that ϕ_h^i are vector valued basis functions, i.e. for e.g. $d = 3$ it is

$$\{\phi_h^i,\ i = 1, \dots, N_h^v\} = \left\{ \left\{ \begin{pmatrix} \psi_h^j \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ \psi_h^j \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ \psi_h^j \end{pmatrix} \right\}, \quad j = 1, \dots, \frac{N_h^v}{3} \right\}$$

5.1 Solution methods for the stationary Stokes problem

We want to find the solution $\{\mathbf{v}_h, p_h\} \in V_h \times L_h \subset H_0^1(\Omega)^d \times L_0^2(\Omega)$ to:

$$(\nabla \mathbf{v}_h, \nabla \phi_h) - (p_h, \nabla \cdot \phi_h) + (\nabla \cdot \mathbf{v}_h, \xi_h) = (\mathbf{f}, \phi_h) \quad \forall \{\phi_h, \xi_h\} \in V_h \times L_h.$$

With (5.1) and (5.2) this is equivalent to the following problem: find $\mathbf{v} \in \mathbb{R}^{N_h^v}$ and $\mathbf{p} \in \mathbb{R}^{N_h^p}$ which solve

$$\begin{aligned} \sum_{j=1}^{N_h^v} (\nabla \phi_h^j, \nabla \phi_h^i) \mathbf{v}_j - \sum_{l=1}^{N_p} (\xi_h^l, \nabla \cdot \phi_h^i) \mathbf{p}_l &= (\mathbf{f}, \phi_h^i) \quad i = 1, \dots, N_h^v, \\ \sum_{j=1}^{N_h^v} (\nabla \cdot \phi_h^j, \xi_h^l) \mathbf{v}_j &= 0 \qquad\qquad l = 1, \dots, N_h^p. \end{aligned} \tag{5.3}$$

This constitutes a system of $N_h^v + N_p$ linear equations and unknowns. We define recall the notation:

$$\mathbf{A} := (\nabla \phi_h^j, \nabla \phi_h^i)_{i,j=1}^{N_h^v \times N_h^v}, \quad \mathbf{B} := -(\nabla \cdot \phi_h^j, \xi_h^i)_{i,j=1}^{N_h^p \times N_h^v}, \quad \mathbf{b} := (\mathbf{f}, \phi_h^i)_{i=1}^{N_h^v} \tag{5.4}$$

leading to the following formulation of the discrete Stokes problem:

$$\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ -\mathbf{B}^T & 0 \end{pmatrix} \begin{pmatrix} \mathbf{v} \\ \mathbf{p} \end{pmatrix} = \begin{pmatrix} \mathbf{b} \\ 0 \end{pmatrix}. \tag{5.5}$$

We recall Lemma 3.3: the matrix $\mathbf{A}$ is symmetric positive definite and the system matrix is positive semidefinite and anti-symmetric in the off-diagonal blocks. We further know that (using an inf-sup stable finite element pair and prescribing a pressure normalization) the discrete Stokes problem is uniquely solvable. The system matrix is regular.

The specific structure of the matrix $\mathbf{A}$ depends on the numbering of the velocity degrees of freedom. Let us assume that $d = 3$ and in the vector $\mathbf{v}^{N_h^v}$ we first have all degrees of freedom of the first velocity component, then all degrees of freedom of the second component and so on. Then, the matrix $\mathbf{A}$ is given as a block diagonal form and the matrix $\mathbf{B}$ can be split component wise as follows:

$$\begin{bmatrix} \mathbf{A}_1 & 0 & 0 & \mathbf{B}_1 \\ 0 & \mathbf{A}_2 & 0 & \mathbf{B}_2 \\ 0 & 0 & \mathbf{A}_3 & \mathbf{B}_3 \\ -\mathbf{B}_1^T & -\mathbf{B}_2^T & -\mathbf{B}_3^T & 0 \end{bmatrix} \begin{pmatrix} \mathbf{v}_1 \\ \mathbf{v}_2 \\ \mathbf{v}_3 \\ \mathbf{p} \end{pmatrix} = \begin{pmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \\ \mathbf{b}_3 \\ 0 \end{pmatrix}, \tag{5.6}$$

where $\mathbf{A}_k = (\nabla\psi_h^j, \nabla\psi_h^i)_{i,j=1}^{N_h^v/3, N_h^v/3}$ are the scalar matrices of the velocity components $k = 1, 2, 3$. The matrices $\mathbf{B}_k$ are given by $\mathbf{B}_k = -(\partial_k\psi_h^i, \xi_h^j)_{i,j=1}^{N_h^v/3, N_h^p}$. The system matrices (5.5) bzw. (5.6) are sparse: the matrices $\mathbf{A}_k$ and $\mathbf{B}_k$ are sparse finite element matrices and there are no coupling terms between the velocity components in the main part for the Stokes equation. However, for the Navier-Stokes equations there is a convection term $\mathbf{v} \cdot \nabla\mathbf{v}$ which will lead to a coupling between the three velocity components.

For $d = 2$, there exist direct highly effective solvers which can optimally exploit the structure of the system matrix. On a modern computer, the Stokes problem with up to $1\,000\,000$ mesh elements can be solved in less than 10min. The storage use of a direct solver, however, grows fast and takes for $1\,000\,000$ mesh elements over 10 GB. For $d = 3$, the solution of the Stokes problem on a mesh with less than $500\,000$ elements takes a few days. The storage use is very large, it takes around 100 GB. Further, in three dimensions more mesh elements are needed than in two dimensions in order to achieve a prescribed accuracy.

For the Navier-Stokes equations, systems of the type (5.6) have to be solved again and again which increases the computational effort.

5.1.1 Schur Complement method

The matrix $\mathbf{A} \in \mathbb{R}^{N_h^v \times N_h^v}$ is symmetric and positiv definite, thus regular. The system (5.5) can symbolically be solved for the pressure. For this, let us multiply the first line with $\mathbf{B}^T\mathbf{A}^{-1}$ from left and add it to the second line leading to

$$\mathbf{B}^T\mathbf{A}^{-1}\mathbf{B}\mathbf{p} = \mathbf{B}^T\mathbf{A}^{-1}\mathbf{b}. \tag{5.7}$$

Thus, we can determine the pressure without knowing the velocity $\mathbf{v}$. If we have determined the pressure, then we can compute the velocity by solving

$$\mathbf{A}\mathbf{v} = \mathbf{b} - \mathbf{B}\mathbf{p}. \tag{5.8}$$

Definition 5.1 (Schur Complement)**.** The matrix

$$\mathbf{\Sigma} := \mathbf{B}^T\mathbf{A}^{-1}\mathbf{B},$$

is called Schur Complement

It holds

Satz 5.2 (Properties of the Schur Complement)**.** The Schur Complement $\mathbf{\Sigma}$ is symmetric and in case of an inf-sup stable discretization positive definite.

Proof. (i) Symmetry follows directly from the structure.

(ii) Positive definite: Let $\mathbf{p} \in \mathbb{R}^{N_h^p}$ be an arbitrary vector. It holds

$$\langle \mathbf{\Sigma}\mathbf{p}, \mathbf{p}\rangle = \langle \mathbf{A}^{-1}\mathbf{B}\mathbf{p}, \mathbf{B}\mathbf{p}\rangle \geqslant c\|\mathbf{B}\mathbf{p}\|^2,$$

due to definiteness of $\mathbf{A}$ if $\mathbf{B}\mathbf{p} \neq 0$. It remains to show that from $\mathbf{B}\mathbf{p} = 0$ it follows $\mathbf{p} = 0$. It holds

$$\mathbf{B}\mathbf{p} = 0 \quad \Rightarrow \quad \langle \mathbf{B}\mathbf{p}, \mathbf{w}\rangle = -(p_h, \nabla \cdot w_h) = 0 \quad \forall w_h \in V_h.$$

According to the inf-sup condition it is

$$\max_{\phi_h \in V_h,\, \|\nabla\phi_h\|=1} (p_h, \nabla \cdot \phi_h) \geqslant \gamma \|p_h\|,$$

thus $p_h = 0$ and therefore $\mathbf{p} = 0$. □Based on (5.7) and (5.8) we can construct two-step strategies in order to compute the solution $\{\mathbf{v}, \mathbf{p}\}$: first we compute the pressure, then we compute the velocity. However, the matrix $\mathbf{\Sigma}$ cannot be assembled directly since the inverse $\mathbf{A}^{-1}$ is usually dense and therefore to expensive to compute. Since both matrices $\mathbf{A}$ and $\mathbf{\Sigma}$ are symmetric and positive definite, efficient iterative solvers can be utilized, e.g. the CG method, which fully rely on matrix-vector products.

With the analogy $\mathbf{A} \sim \Delta^{-1}$, $\mathbf{B} \sim$ grad und $\mathbf{B}^T \sim$ div we obtain for the Schur Complement the heuristic approximation by counting the derivative orders:

$$\mathbf{\Sigma} \sim \text{div} \circ \Delta^{-1} \circ \text{grad} \sim \text{id}\,.$$

The Schur Complement matrix therefore roughly behaves as an operator of zero order, in the finite element context as the mass matrix $\mathbf{M}_p$ in the pressure space:

$$\mathbf{M}_p = (\xi_h^j, \xi_h^i)_{i,j=1}^{N_h^p, N_h^p}.$$

The condition of the matrix behaves on regular grids as $\text{cond}_2(\mathbf{M}_p) = O(1)$.

Satz 5.3 (Condition of the mass matrix)**.** Let $\mathbf{M}$ denote the mass matrix associated with a regular finite element mesh. Then it is

$$\mathrm{cond}_2(\mathbf{M}) = O(1).$$

Proof. Let us denote the nodal basis functions as ξ_h^i with $i = 1, \dots, N_h^p$ and the finite element space as L_h. Due to the regularity of the grid and the norm equivalence, it holds with a constant $c > 0$ uniformly in h:

$$\frac{h}{c} = \frac{h}{c}\|\xi_h^i\|_\infty \leqslant \|\xi_h^i\| \leqslant ch\|\xi_h^i\|_\infty = ch.$$

From this, it follows for $p_h \in L_h$ with coefficient vector $\mathbf{p} \in \mathbb{R}^{N_h^p}$:

$$\langle \mathbf{M}\mathbf{p}, \mathbf{p}\rangle = \|p_h\|^2 \leqslant \sum_K \|p_h\|_K^2 \leqslant \sum_K \| \sum_{x_i \in K} \mathbf{p}_i \xi_h^i\|^2 \leqslant c_K \sum_i \mathbf{p}_i^2 \|\xi_h^i\|^2 \leqslant c c_K h^2 \langle \mathbf{p}, \mathbf{p}\rangle,$$

where the constant $c_K > 0$ depends on how the finite element basis overlap. With the Rayleigh quotient it follows for the largest eigenvalue

$$\lambda_{\max}(\mathbf{M}) = \max_{\mathbf{p}} \frac{\langle \mathbf{M}\mathbf{p}, \mathbf{p}\rangle}{\langle \mathbf{p}, \mathbf{p}\rangle} \leqslant c c_K h^2.$$

Accordingly, we have for the smallest eigenvalue:

$$\lambda_{\min}(\mathbf{M}) = \min_{\mathbf{p}} \frac{\langle \mathbf{M}\mathbf{p}, \mathbf{p}\rangle}{\langle \mathbf{p}, \mathbf{p}\rangle} \geqslant \frac{c_K}{c} c h^2,$$

and it follows $\mathrm{cond}_2(\mathbf{M}) = O(1)$. □As crucial argument in the proof, we used the regularity of the mesh. On locally refined or e.g. strongly anisotropic meshes this derivation for the condition number of the mass matrix does not hold anymore.

Due to the heuristic estimation $\mathbf{\Sigma} \sim \mathbf{M}_p$ we will derive a similar result for the Schur Complement. We will not analyze the eigenvalues of the Schur Complement, but the eigenvalues of the Schur Complement preconditioned with the mass matrix: $\mathbf{M}_p^{-1}\mathbf{\Sigma}$. By this we can reduce the dependency on the regularity of the mesh.

Theorem 5.4 (Condition of the Schur Complement matrix)**.** For the Schur Complement $\mathbf{\Sigma} := \mathbf{B}^T\mathbf{A}^{-1}\mathbf{B}$ and the mass matrix $\mathbf{M}_p$ in the pressure space L_h it holds:

$$\mathrm{cond}_2(\mathbf{M}^{-1}\mathbf{\Sigma}) \leqslant c_0\gamma^{-2}, \quad \|\mathbf{M}_p^{-1}\mathbf{\Sigma}\| \leqslant c_0,$$

where γ is the inf-sup constant for the finite element pair $V_h \times L_h$ and $c_0 \leqslant 4$ is a constant.

Proof. (i) Let us analyze the eigenvalues of $\mathbf{M}_p^{-1}\mathbf{\Sigma}$. Let $\mathbf{q}$ be an eigenvector and λ an eigenvalue with:

$$\mathbf{M}_p^{-1}\mathbf{\Sigma}\mathbf{q} = \lambda\mathbf{q} \quad \Leftrightarrow \quad \lambda = \frac{\langle \mathbf{\Sigma}\mathbf{q}, \mathbf{q}\rangle}{\langle \mathbf{M}_p\mathbf{q}, \mathbf{q}\rangle} = \frac{\langle \mathbf{A}^{-1}\mathbf{B}\mathbf{q}, \mathbf{B}\mathbf{q}\rangle}{\langle \mathbf{M}_p\mathbf{q}, \mathbf{q}\rangle} =: \frac{\langle \mathbf{B}\mathbf{q}, \mathbf{B}\mathbf{q}\rangle^2_{\mathbf{A}^{-1}}}{\langle \mathbf{M}_p\mathbf{q}, \mathbf{q}\rangle}. \tag{5.9}$$

Since $\mathbf{A}$ is symmetric and positiv definite, $\mathbf{A}^{-1}$ defines an inner product $\langle\cdot,\cdot\rangle_{\mathbf{A}^{-1}}$ with norm $|\cdot|_{\mathbf{A}^{-1}} = \langle\cdot,\cdot\rangle_{\mathbf{A}^{-1}}^{\frac{1}{2}}$. We can also write this norm as

$$|\mathbf{x}|_{\mathbf{A}^{-1}} := \max_{\mathbf{y}\in\mathbb{R}^{N_h^v}} \frac{\langle \mathbf{x},\mathbf{y}\rangle_{\mathbf{A}^{-1}}}{|\mathbf{y}|_{\mathbf{A}^{-1}}} = \max_{\mathbf{y}\in\mathbb{R}^{N_h^v}} \frac{\langle \mathbf{A}^{-1}\mathbf{x},\mathbf{y}\rangle}{\langle \mathbf{A}^{-1}\mathbf{y},\mathbf{y}\rangle^{\frac{1}{2}}}. \tag{5.10}$$

(*ii*) Using (5.9) and (5.10) we obtain with $\mathbf{z} := \mathbf{A}^{-1}\mathbf{y}$:

$$\lambda = \max_{\mathbf{y}\in\mathbb{R}^{N_v}} \frac{\langle \mathbf{A}^{-1}\mathbf{B}\mathbf{q},\mathbf{y}\rangle^2}{\langle \mathbf{M}_p\mathbf{q},\mathbf{q}\rangle\langle \mathbf{A}^{-1}\mathbf{y},\mathbf{y}\rangle} = \max_{\mathbf{z}\in\mathbb{R}^{N_v}} \frac{\langle \mathbf{B}\mathbf{q},\mathbf{z}\rangle^2}{\langle \mathbf{M}_p\mathbf{q},\mathbf{q}\rangle\langle \mathbf{A}\mathbf{z},\mathbf{z}\rangle}$$

Let $\mathbf{z}_h \in V_h$ a finite element function with coefficient vector $\mathbf{z} \in \mathbb{R}^{N_h^v}$ and let $q_h \in L_h$ be a finite element function with coefficient vector $\mathbf{q} \in \mathbb{R}^{N_h^p}$. It follows:

$$\lambda = \max_{\mathbf{z}_h\in V_h} \frac{(q_h, \nabla\cdot\mathbf{z}_h)^2}{(q_h,q_h)(\nabla\mathbf{z}_h,\nabla\mathbf{z}_h)}$$

(*iii*) For the smallest eigenvalue we get with the inf-sup condition:

$$\lambda_{\min}(\mathbf{M}_p^{-1}\mathbf{\Sigma}) = \min_{q_h\in L_h}\max_{\mathbf{z}_h\in V_h} \frac{(q_h, \nabla\cdot\mathbf{z}_h)^2}{(q_h,q_h)(\nabla\mathbf{z}_h,\nabla\mathbf{z}_h)} \geqslant \gamma_h^2.$$

(*iv*) For the largest eigenvalue we obtain

$$\lambda_{\max}(\mathbf{M}_p^{-1}\mathbf{\Sigma}) = \max_{q_h\in L_h}\max_{\mathbf{z}_h\in V_h} \frac{(q_h, \nabla\cdot\mathbf{z}_h)^2}{(q_h,q_h)(\nabla\mathbf{z}_h,\nabla\mathbf{z}_h)} \leqslant \max_{q_h\in L_h}\max_{\mathbf{z}_h\in V_h} \frac{\|q_h\|^2\,\|\nabla\cdot\mathbf{z}_h\|^2}{\|q_h\|^2\,\|\nabla\mathbf{z}_h\|^2}.$$

For H^1 conformal finite element spaces $V_h \subset H_0^1(\Omega)^d$ it holds $\|\nabla\cdot\mathbf{z}_h\| \leqslant \|\nabla\mathbf{z}_h\|$ (compare Lemma 2.6). If $d = 3$ it holds $\|\nabla\cdot\mathbf{z}_h\|^2 \leqslant 4\|\nabla\mathbf{z}_h\|^2$ leading to $\lambda_{\max}(\mathbf{M}_p^{-1}\mathbf{\Sigma}) \leqslant 4$.

□

The Schur Comlement matrix preconditioned with the mass matrix is therefore well conditioned. In particular, the condition is independent of the mesh size h. For this reason, we consider the following solution method for preconditioned systems:

$$\mathbf{M}_p^{-1}\mathbf{\Sigma}\mathbf{p} = \mathbf{M}_p^{-1}\mathbf{B}^T\mathbf{A}^{-1}\mathbf{b}. \tag{5.11}$$

The Uzawa algorithm A classical method for the solution of the Stokes system is the *Uzawa algorithm*. The basic idea is to approximate the preconditioned Schur Complement problem (5.11) with a *damped Richardson iteration*. Let $\mathbf{p}^{(0)} \in \mathbb{R}^{N_h^p}$ be an initial value. We consider the iteration:

$$\begin{aligned}\mathbf{p}^{(t+1)} &= \mathbf{p}^{(t)} + \omega(\mathbf{M}_p^{-1}\mathbf{B}^T\mathbf{A}^{-1}\mathbf{b} - \mathbf{M}_p^{-1}\mathbf{\Sigma}\mathbf{p}^{(t)})\\ &= \mathbf{p}^{(t)} + \omega\mathbf{M}_p^{-1}(\mathbf{B}^T\mathbf{A}^{-1}\mathbf{b} - \mathbf{B}^T\mathbf{A}^{-1}\mathbf{B}\mathbf{p}^{(t)})\\ &= \mathbf{p}^{(t)} + \omega\mathbf{M}_p^{-1}\mathbf{B}^T\mathbf{A}^{-1}(\mathbf{b} - \mathbf{B}\mathbf{p}^{(t)}),\end{aligned} \tag{5.12}$$

where $\omega > 0$ is a suitable damping relaxation parameter. The method is formulated as a two-step iteration and delivers an approximation for the velocity $\mathbf{v} \in \mathbb{R}^{N_h^v}$ at the same time.

Algorithmus 5.5 (The Uzawa algorithm)**.**

$$\begin{aligned} \text{Initial value}: \quad & \mathbf{p}^{(0)} \in \mathbb{R}^{N_p} \\ \text{for } t \geqslant 0: \quad & \mathbf{A}\mathbf{v}^{(t)} = \mathbf{b} - \mathbf{B}\mathbf{p}^{(t)} \\ & \mathbf{M}_p \mathbf{p}^{(t+1)} = \mathbf{M}_p \mathbf{p}^{(t)} + \omega \mathbf{B}^T \mathbf{v}^{(t)}, \end{aligned}$$

with $\omega > 0$ relaxation parameter.

In order to analyze the convergence of the Uzawa algorithm, we need the following auxiliary result.

Lemma 5.6. Let $\mathbf{A}$ be a symmetric and positive definite matrix and let $q \in (0,1)$ be given. Let for the eigenvalues of $\mathbf{A}$ hold true:

$$\lambda(\mathbf{A}) \in [q, 2-q].$$

Then, it follows that

$$\|\mathbf{I} - \mathbf{A}\|_2 \leqslant 1 - q.$$

Proof. Since $\mathbf{A}$ is symmetric, it holds for the spectral norm

$$\|\mathbf{I} - \mathbf{A}\|_2 = \max_{x \in \mathbb{R}^n} \left| \frac{\langle \mathbf{x} - \mathbf{A}\mathbf{x}, \mathbf{x} \rangle}{\langle \mathbf{x}, \mathbf{x} \rangle} \right| = \max_{x \in \mathbb{R}^n} \left| 1 - \frac{\langle \mathbf{A}\mathbf{x}, \mathbf{x} \rangle}{\langle \mathbf{x}, \mathbf{x} \rangle} \right|.$$

All eigenvalues of $\mathbf{A}$ are positive and real, thus $0 < \lambda_{min}(\mathbf{A}) \leqslant \lambda_{max}(\mathbf{A})$:

$$\|\mathbf{I} - \mathbf{A}\|_2 = \max\{|1 - \lambda_{max}(\mathbf{A})|, |1 - \lambda_{min}(\mathbf{A})|\}.$$

This means that for $\{\lambda_{max}(\mathbf{A}), \lambda_{min}(\mathbf{A})\} \in [q, 2-q]$ the claim holds. □

The following result describes the convergence behavior of the Uzawa iteration:

Theorem 5.7 (Convergence of the Uzawa algorithm)**.** For every $\omega < \frac{1}{4}$ the Uzawa iteration converges to the solution $\{\mathbf{v}, \mathbf{p}\}$ of the saddle point problem (5.5) with the estimation:

$$|\mathbf{p}^{(t)} - \mathbf{p}| \leqslant \rho^t |\mathbf{p}^{(0)} - \mathbf{p}|$$

and the convergence rate $\rho \leqslant 1 - \omega\gamma^2$, where γ^2 is the inf-sup constant of the ansatz space.

Proof. With the help of the iteration, we derive a connection between the old and the new pressure error. With (5.12) and (5.11) it holds

$$\begin{aligned} \mathbf{p} - \mathbf{p}^{(t+1)} &= \mathbf{p} - \mathbf{p}^{(t)} - \omega \mathbf{M}_p^{-1} (\mathbf{B}^T \mathbf{A}^{-1} \mathbf{p} - \mathbf{\Sigma} \mathbf{p}^{(t)}) \\ &= \mathbf{p} - \mathbf{p}^{(t)} - \omega \mathbf{M}_p^{-1} \mathbf{\Sigma} (\mathbf{p} - \mathbf{p}^{(t)}) \\ &= [\mathbf{I} - \omega \mathbf{M}_p^{-1} \mathbf{\Sigma}] (\mathbf{p} - \mathbf{p}^{(t)}), \end{aligned}$$

thus

$$|\mathbf{p} - \mathbf{p}^{(t+1)}| \leqslant \|I - \omega \mathbf{M}_p^{-1}\mathbf{\Sigma}\| \, |\mathbf{p} - \mathbf{p}^{(t)}|.$$

We now would like to apply Lemma 5.6 to the matrix $\omega \mathbf{M}_p^{-1}\mathbf{\Sigma}$. However, this matrix is not symmetric but has the same eigenvalues as the symmetric matrix $\mathbf{M}_p^{-\frac{1}{2}}\mathbf{\Sigma}\mathbf{M}_p^{-\frac{1}{2}}$, where we can apply the Lemma 5.6. For the eigenvalues of the matrix $\mathbf{M}_p^{-1}\mathbf{\Sigma}$ we obtain by Theorem 5.4:

$$\lambda_{\min}(\mathbf{M}_p^{-1}\mathbf{\Sigma}) \geqslant \gamma^2, \quad \lambda_{\max}(\mathbf{M}_p^{-1}\mathbf{\Sigma}) \leqslant c_0,$$

with the inf-sup constant $\gamma > 0$ and wiht $c_0 \leqslant 4$. For $\omega' = \frac{1}{4}$ it holds:

$$\lambda_{\min}(\omega' \mathbf{M}_p^{-1}\mathbf{\Sigma}) \geqslant \frac{\gamma^2}{4}, \quad \lambda_{\max}(\omega' \mathbf{M}_p^{-1}\mathbf{\Sigma}) \leqslant 1,$$

and Lemma 5.6 delivers the convergence rate

$$\rho \leqslant 1 - \frac{1}{4}\gamma^2.$$

The estimation follows by applying the iteration formula again and again. □

The convergence rate of the Uzawa iteration depends on the inf-sup constant. For this reason, we can use iteration methods for the inversion of the Schur Complement in order to experimentally determine the inf-sup constant. It can be shown that the Uzawa algorithm is equivalent to the application of a gradient method to the Schur Complement system. Since the Schur Complement is symmetric, we can use the CG method for an approximation. We use the mass matrix $\mathbf{M}_p$ as preconditioner and can expect optimal convergence order $\mathrm{cond}_2(\mathbf{M}_p^{-1}\mathbf{\Sigma}) = O(1)$ according to Theorem 5.4.

Das PCG-Verfahren Wir betrachten nun das mit der Massematrix $\mathbf{M}_p$ vorkonditionierte PCG-Verfahren (*preconditioned conjugate gradients*), angewendet auf das Schur-Komplement System (5.11):

Algorithmus 5.8 (PCG-Iteration f"ur das Schur-Komplement)**.**

$$\begin{aligned}
\text{Startwert}: \quad & \mathbf{p}^{(0)} \in \mathbb{R}^{N_p}, \quad \mathbf{r}^{(0)} := \mathbf{B}^T\mathbf{A}^{-1}\mathbf{b} - \mathbf{\Sigma}\mathbf{p}^{(0)}, \quad \mathbf{z}^{(0)} := \mathbf{M}_p^{-1}\mathbf{r}^{(0)}, \quad \mathbf{q}^{(0)} := \mathbf{z}^{(0)} \\
\text{f"ur } t \geqslant 0: \quad & \alpha_t := \frac{\langle \mathbf{r}^{(t)}, \mathbf{z}^{(t)} \rangle}{\langle \mathbf{\Sigma}\mathbf{q}^{(t)}, \mathbf{q}^{(t)} \rangle}, \\
& \mathbf{p}^{(t+1)} := \mathbf{p}^{(t)} + \alpha_t \mathbf{q}^{(t)}, \quad \mathbf{r}^{(t+1)} := \mathbf{r}^{(t)} - \alpha_t \mathbf{\Sigma}\mathbf{q}^{(t)}, \\
& \mathbf{z}^{(t+1)} := \mathbf{M}_p^{-1}\mathbf{r}^{(t+1)}, \\
& \beta_t := \frac{\langle \mathbf{r}^{(t+1)}, \mathbf{z}^{(t+1)} \rangle}{\langle \mathbf{r}^{(j)}, \mathbf{z}^{(j)} \rangle}, \\
& \mathbf{q}^{(t+1)} := \mathbf{z}^{(t+1)} + \beta \mathbf{q}^{(t)}.
\end{aligned}$$

Der PCG-Algorithmus erfordert in jedem Schritt eine Matrix-Vektor Multiplikation mit dem Schur-Komplement $\mathbf{\Sigma}$. Da $\mathbf{\Sigma}$ nicht explizit berechnet werden kann muss hierf"ur jeweils die Matrix $\mathbf{A}$ invertiert werden. Diese Matrix hat allerdings die sehr einfache Gestalt einer Block-Laplace-Matrix und die Invertierung kann z.B. effizient mit Mehrgitterverfahren erfolgen. Desweiteren ist in jedem Schritt die Invertierung der Massematrix $\mathbf{M}_p$ notwendig. Aufgrund der guten Konditionszahl kann dies z.B. wieder mit dem CG-Verfahren geschehen. Das vorkonditionierte CG-Verfahren angewendet auf (5.11) konvergiert mit der Absch"atzung:

$$|\mathbf{p}^{(t)} - \mathbf{p}| \leqslant \kappa \left(\frac{1 - \kappa^{-\frac{1}{2}}}{1 + \kappa^{-\frac{1}{2}}} \right)^t |\mathbf{p}^{(0)} - \mathbf{p}|, \quad \kappa := \mathrm{cond}_2(\mathbf{M}_p^{-1}\mathbf{\Sigma}).$$

In jedem Schritt wird also eine feste Konvergenzrate erzielt, welche im Wesentlichen von der Konstante γ der diskreten *inf-sup Bedingung* abh"angt. Der Aufwand wird durch die Multiplikation mit $\mathbf{\Sigma}$ bestimmt. Gelingt es, die Matrix $\mathbf{A}$ mit optimalem Aufwand zu invertieren ($O(N)$), also z.B. mit einem Mehrgitterverfahren, so besitzt auch das vorkonditionierte CG-Verfahren die optimale Komplexit"at $O(N)$ zur Reduktion des Fehlers um einen konstanten Faktor.

5.1.2 L"osung der Laplace-Matrix

Sowohl beim CG-Verfahren als bei der Uzawa-Iteration muss in jedem Schritt die Geschwindigkeitsmatrix $\mathbf{A}$ zu gegebener rechter Seite $\mathbf{Ax} = \mathbf{b}$ invertiert werden. Aufgrund der speziellen Matrix-Struktur (5.6) kann dies f"ur jede Komponente getrennt geschehen:

$$\mathbf{A}_d \mathbf{x}_d = \mathbf{b}_d.$$

Dabei unterscheiden sich die Matrizen $\mathbf{A}_d$ nur, wenn verschiedene Randwerte vorgegeben sind, z.B. bei der Verwendung der *slip-Bedingung* $\mathbf{v} \cdot \mathbf{n} = 0$ (dann gelten Dirichlet-Werte nur f"ur die Normalkomponente von $\mathbf{v}$). Im Wesentlichen ist also die Laplace-Matrix im Geschwindigkeitsraum zu invertieren. Dies kann mit Mehrgitterverfahren in optimaler Komplexit"at $O(N_v)$ erreicht werden.

5.1.3 L"osung von stabilisierten Systemen

Bei Finite-Elemente-Paaren $Q_h \times V_h$ welche nicht *inf-sup stabil* sind, werden der Bilinearform zus"atzliche Stabilisierungsterme hinzugef"ugt, siehe Kapitel 3.3, z.B.:

$$(\nabla \mathbf{v}_h, \nabla \phi_h) - (p_h, \nabla \cdot \phi_h) + (\nabla \cdot \mathbf{v}_h, \xi_h) + \sum_{K \in \Omega_h} h_K^2 (\nabla p_h, \nabla \xi_h)_K = (\mathbf{f}, \phi_h).$$

Das Stokes-System in Sattelpunktform ist dann anstelle von (5.5)

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ -\mathbf{B}^T & \mathbf{S} \end{bmatrix} \begin{bmatrix} \mathbf{v} \\ \mathbf{p} \end{bmatrix} = \begin{bmatrix} \mathbf{b} \\ 0 \end{bmatrix}, \tag{5.13}$$

mit der Stabilisierungsmatrix $\mathbf{S}$ und das modifizierte Schurkomplement $\mathbf{\Sigma}_S$ hat die Form

$$\mathbf{\Sigma}_S := \mathbf{S} + \mathbf{B}^T\mathbf{A}^{-1}\mathbf{B}.$$

Auch diese Matrix ist symmetrisch positiv definit, so dass Uzawa, bzw. CG-Verfahren entsprechend angewendet werden k"onnen.

5.2 L"osung des station"aren Navier-Stokes-Problem

Die Navier-Stokes-Gleichungen unterscheiden sich von den Stokes-Gleichungen durch den zus"atzlichen Konvektionsterm $(\mathbf{v}\cdot\nabla\mathbf{v},\phi)$. Wir gehen davon aus, dass entsprechend Kapitel 3.4.1 die L"osung mit einem Iterationsverfahren, z.B. dem Newton-Verfahren gewonnen wird. Sei $\tilde{\mathbf{v}}$ die letzte Approximation. Dann ist die diskrete L"osung $(\mathbf{p},\mathbf{v})$ durch ein Sattelpunktproblem gegeben:

$$\begin{bmatrix}\mathbf{A}(\tilde{\mathbf{v}}) & \mathbf{B}\\ -\mathbf{B}^T & 0\end{bmatrix}\begin{bmatrix}\mathbf{v}\\ \mathbf{p}\end{bmatrix} = \begin{bmatrix}\mathbf{b}\\ 0\end{bmatrix}. \tag{5.14}$$

Die genaue Form der Matrix $A(\tilde{\mathbf{v}})$ h"angt vom gew"ahlten Linearisierungsverfahren ab:

$$\text{Oseen:}\quad \mathbf{A}(\tilde{\mathbf{v}}) = \left((\nu\nabla\phi_h^j,\nabla\phi_h^i) + (\tilde{\mathbf{v}}_h\cdot\nabla\phi_h^j,\phi_h^i)\right)_{i,j=1}^{N_v}$$

$$\text{Newton:}\quad A(\tilde{\mathbf{v}}) = \left((\nu\nabla\phi_h^j,\nabla\phi_h^i) + (\tilde{\mathbf{v}}_h\cdot\nabla\phi_h^j,\phi_h^i) + (\phi_h^j\cdot\nabla\tilde{\mathbf{v}}_h,\phi_h^i)\right)_{i,j=1}^{N_v}.$$

Die Matrix $\mathbf{A}(\tilde{\mathbf{v}})$ ist in beiden F"allen nicht symmetrisch. Im Fall der Newton-Linearisierung weist die Matrix $A(\tilde{\mathbf{v}})$ auch keine Diagonal-Blockstruktur auf. Anstelle des vereinfachten Vorgehens in Abschnitt 5.1.2 muss also stets das gekoppelte Geschwindigkeitssystem $A(\tilde{\mathbf{v}})\mathbf{v} = \mathbf{b} - \mathbf{B}\mathbf{p}$ gel"ost werden. F"ur kleine Reynoldszahlen ist dies mit Mehrgitterverfahren in optimaler Komplexit"at m"oglich. F"ur gro"se Reynoldszahlen ist die Konstruktion von robusten Gl"attungsverfahren schwierig.

5.2.1 Schur-Komplement Methoden

Wie das Stokes-System kann das diskrete Navier-Stokes Problem mit Schur-Komplement-Verfahren gel"ost werden:

$$\mathbf{B}^T\mathbf{A}(\tilde{\mathbf{v}})^{-1}\mathbf{B}\mathbf{p} = \mathbf{B}^T\mathbf{A}(\tilde{\mathbf{v}})^{-1}\mathbf{b}. \tag{5.15}$$

Anstelle der einfachen Laplace-Matrix muss nun in jedem Schritt die Matrix $A(\tilde{\mathbf{v}})$ invertiert werden. Diese Matrix ist nicht symmetrisch (aufgrund der Nichtlinearit"at), daher kann das Schur-Komplement nicht mit dem CG-Verfahren invertiert werden. Als Alternative k"onnen jedoch andere Krylow-Raum-Verfahren, wie *GMRES* oder *BiCGStab* verwendet werden, welche auch auf nicht-symmetrische Probleme angewendet werden k"onnen. Die Invertierung dieser Matrix ist bei weitem aufw"andiger als die Invertierung der Laplace-Matrix beim Stokes-System. F"ur die station"aren Navier-Stokes-Gleichungen werden daher oft L"osungsmethoden verwendet, welche unmittelbar das gekoppelte Sattelpunktsystem zu l"osen versuchen.

5.2.2 Mehrgitterverfahren

Das System (5.14) ist regul"ar (denn es existiert eine eindeutige L"osung), jedoch nicht symmetrisch oder positiv definit. Das System kann z.B. mit einem (vorkonditionierten) Krylowraum-Verfahren (GMRES) gel"ost werden. Zur Vereinfachung f"uhren wir die folgende Bezeichnung ein:

$$\mathbf{X}\mathbf{u} = \mathbf{f}, \quad \mathbf{u} = (\mathbf{p}, \mathbf{v})^T, \quad \mathbf{f} = (\mathbf{b}, 0)^T.$$

Die Matrix $\mathbf{X}(\tilde{\mathbf{u}})$ ist schlecht konditioniert, es gilt im Allgemeinen $\mathrm{cond}_2(\mathbf{X}) = O(\nu^{-1}h^{-2}\gamma^{-2})$, mit der Konstante der *inf-sup Bedingung* γ und der Viskosit"at ν. Zum L"osen des Systems, oder aber als Vorkonditionierer im GMRES-Verfahren bietet sich somit ein Mehrgitter-Verfahren an. Wir rekapitulieren hier kurz die Idee: Einfache Iterationsverfahren wie das Jacobi- oder Gau"s-Seidel-Verfahren d"ampfen sehr schnell *hochfrequente* Fehleranteile. Im Mehrgitterkontext wird diese Iteration deshalb *Gl"attungsoperator* genannt und hier mit $\mathcal{S}$ bezeichnet. Bei der Finite-Elemente Approximation sind dies Fehleranteile $\mathbf{v} - \mathbf{v}_h$, welche Oszillationen auf der feinsten Gitterebene h darstellen. Schwingungen mit niedriger Frequenz hingegen werden nur sehr langsam reduziert. Die Idee ist jetzt, auf einer Gitterebene nur die *hochfrequenten* Anteile zu *gl"atten*, und die niederfrequenten Fehleranteile auf einem gr"oberen Gitter mit Gitterweite $2h$ zu behandeln. Dieses Verfahren kann iterativ angewendet werden:

Algorithmus 5.9 (Mehrgitter-Algorithmus $\mathbf{X}_h\mathbf{u}_h = \mathbf{f}_h$ auf Gitter Ω_h)**.** Ist $\Omega_h = \Omega_0$ das Grobgitter, so l"ose $\mathbf{X}_0\mathbf{u}_0 = \mathbf{f}_0$ mit einem direkten L"oser.

1) Vorgl"atten:	$\mathbf{u}_h^1$	$:= \mathcal{S}(\mathbf{u}_h, \mathbf{f}_h, \mathbf{X}_h)$,
2) Defekt:	$\mathbf{d}_h$	$:= \mathbf{f}_h - \mathbf{X}_h\mathbf{u}_h^1$,
3) Restriktion:	$\mathbf{d}_H$	$:= \mathcal{R}_H\mathbf{d}_h$,
4) Grobgritterl"osung:	$\mathbf{X}_H\mathbf{y}_H$	$:= \mathbf{d}_H$,
5) Prolongation:	$\mathbf{y}_h$	$:= \mathcal{P}_h\mathbf{y}_H$,
6) Update:	$\mathbf{u}_h^2$	$:= \mathbf{u}_h^1 + \theta\mathbf{y}_h$,
7) Nachgl"atten:	$\mathbf{u}_h^3$	$:= \mathcal{S}(\mathbf{u}_h^2, \mathbf{f}_h, \mathbf{X}_h)$.

Die *Grobgitterl"osung* in Schritt *4*) erfolgt rekursiv durch das Mehrgitterverfahren selbst. Lediglich auf dem gr"obsten Gitter Ω_0 wird das Problem exakt gel"ost (oder hinreichend genau mit einem iterativen Verfahren). Das Mehrgitter-Verfahren beruht auf einer *Hierarchie von Gittern* $\Omega_H = \Omega_0,\ \Omega_1, \ldots, \Omega_L = \Omega_h$. Auf jedem dieser Gitter Ω_l wird der Finite Elemente Raum $Q_l \times V_l$ und somit das Problem $\mathbf{X}_l\mathbf{u}_l = \mathbf{f}_l$ definiert. Um die optimale Komplexit"at $O(N)$ zu erreichen, m"ussen die Gl"attungsoperationen $\mathcal{S}$ in Schritt *1*) und *7*) alle *hochfrequenten* Anteile des Fehlers mit einer festen (von der Gitterweite h unabh"angigen) Rate *gl"atten*. Dann ist zur Reduktion des Fehlers um einen festen Faktor auf jeder Gitterebene eine feste Anzahl von Schritten durchzuf"uhren. Der L"osungsschritt auf dem Grobgitter verschlechtert die Komplexit"at nicht, da $\#\Omega_0 = O(1)$ angenommen werden kann. Solche geometrische Mehrgitterverfahren k"onnen optimal implementiert werden

und erreichen die theoretische Komplexit"at $O(N)$, allerdings mit einer sehr gro"sen Konstante. Bis zu einigen Tausend Elementen sind direkte Verfahren zur L"osung der Navier-Stokes Gleichungen effizienter.

Das Mehrgitter-Verfahren kann entweder als L"oser innerhalb einer Schur-Komplement-Iteration f"ur die Probleme (5.11) oder (5.15) verwendet werden oder unmittelbar auf das Sattelpunktproblem (5.14) angewendet werden.

Mehrgitter-Gl"atter in der Schur-Komplement-Iteration Wird z.B. das CG-Verfahren verwendet um das Schur-Komplement-System der Stokes-Gleichungen zu l"osen, so muss zur Matrix-Vektor-Multiplikation mit dem Schur-Komplement $\mathbf{\Sigma} := \mathbf{B}^T\mathbf{A}^{-1}\mathbf{B}$ die Matrix $\mathbf{A}$ *invertiert* werden:

$$\mathbf{A}\mathbf{x} = \mathbf{b},$$

wobei $\mathbf{A}$ eine Block-Laplace-Matrix ist, also von sehr einfache Typ. Die rechte Seite $\mathbf{b}$ ist das Residuum in der Schur-Komplement Iteration. Die Kondition von $\mathbf{A}$ h"angt von der Gitterweite per $\mathrm{cond}_2(\mathbf{A}) = O(h^{-2})$ ab. F"ur die Laplace-Gleichung existieren sehr einfache Mehrgitterverfahren. Schon einige Schritte des *ged"ampften Jacobi-Verfahrens* als Gl"atter liefern optimale Komplexit"at und Konvergenz:

$$\mathcal{S}(\mathbf{x}_h, \mathbf{b}_h, \mathbf{A}_h) = \mathbf{x}_h + \theta \mathrm{diag}(\mathbf{A}_h)^{-1}(\mathbf{b}_h - \mathbf{A}_h\mathbf{x}_h), \quad \theta < \frac{1}{2}.$$

Bei der Schur-Komplement-Iteration f"ur die Navier-Stokes Gleichungen muss in jedem Schritt die nicht symmetrische Matrix

$$\mathbf{A}(\tilde{\mathbf{v}})\mathbf{x} = \mathbf{b},$$

invertiert werden. F"ur gr"o"sere Reynoldszahlen, wenn das Problem also stark konvektionsdominant ist, so ist die Konstruktion von robusten Gl"attern schwierig. F"ur einen gro"sen Bereich ist eine unvollst"andige LU-Zerlegung erfolgreich:

$$\mathcal{S}(\mathbf{x}_h, \mathbf{b}_h, \mathbf{A}_h) = \mathbf{x}_h + \theta \mathcal{ILU}(\mathbf{A}_h(\tilde{\mathbf{v}}))^{-1}(\mathbf{b}_h - \mathbf{A}_h\mathbf{x}_h).$$

Die ILU (incomplete lower upper) Zerlegung ist dabei eine Zerlegung von $\mathbf{A}_h(\tilde{\mathbf{v}})$ in eine linke untere und rechte obere Dreiecksmatrix. "Ublicherweise ist die LU-Zerlegung einer d"unn besetzten Matrix voll besetzt. Bei der ILU-Zerlegung wird die Besetzungsstruktur beibehalten. Der Erfolg der ILU-Zerlegung als Gl"atter h"angt stark von der Sortierung der Freiheitsgrade in der Matrix ab.

Mehrgitter-Gl"atter f"ur das Sattelpunktproblem Soll unmittelbar das Sattelpunktproblem

$$\begin{bmatrix} \mathbf{A}(\tilde{\mathbf{v}}) & \mathbf{B} \\ -\mathbf{B}^T & 0 \end{bmatrix} \mathbf{x} = \mathbf{b},$$

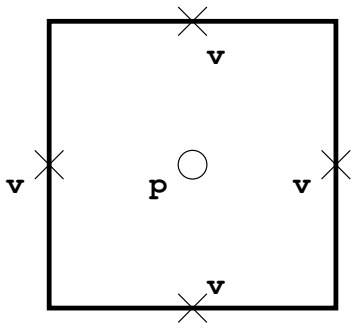


Figure 5.1: Die 9 Freiheitsgrade ($4 \cdot 2$ Geschwindigkeiten, 1 Druck) des rotierten $\tilde{Q}^1 - P^0$ Elements in zwei Dimensionen.

mit dem Mehrgitter-Verfahren gel"ost werden, so ist die Wahl des Gl"atters schwierig. Die einfachen Iterationsverfahren versagen. Erfolg verspricht eine *Blockung* von benachbarten Freiheitsgraden. Im Folgenden stellen wir zwei Methoden vor.

Ein klassischer Gl"atter f"ur Sattelpunktprobleme ist der *Vanka-Gl"atter*. Wir entwickeln das Verfahren f"ur das rotierte $\tilde{Q}^1 - P^0$ Element (Abbildung 5.1). F"ur eine Zelle $K \in \Omega_h$ der Triangulierung sei durch

$$\mathbf{X}_K = \begin{bmatrix} \mathbf{A}(\tilde{\mathbf{v}})\big|_K & \mathbf{B}\big|_K \\ -\mathbf{B}^T\big|_K & 0 \end{bmatrix}, \quad \mathbf{x}_K := \mathbf{x}\big|_K, \quad \mathbf{b}_K := \mathbf{b}\big|_K,$$

die Einschr"ankung der Matrix, der L"osung und der rechten Seite auf die Freiheitsgrade der Zelle K gegeben. Die Matrix $\mathbf{X}_K \in \mathbb{R}^{9\times 9}$ ist klein und kann exakt invertiert werden. Der Vanka-Gl"atter wird als eine Block-Gau"s-Seidel Iteration aufgebaut. Der Gl"atter l"auft in einer Iteration "uber alle Zellen $K_i \in \Omega_i$:

$$i = 1, 2, \cdots : \quad \mathbf{X}_{K_i}\tilde{\mathbf{x}}_{K_i} = \mathbf{b}_{K_i}(K_1, \ldots, K_{i-1}).$$

Bereits berechnete Gr"o"sen $\tilde{\mathbf{x}}_{K_i}$ werden in der rechten Seite ber"ucksichtigt (ansonsten w"are das Verfahren eine Block-Jacobi-Iteration). Zur Erh"ohung der Robustheit wird ein D"ampfungsparameter $\theta \in (0, 1)$ eingef"uhrt:

$$\mathbf{x}^{i+1} = \mathbf{x}^i + \theta(\tilde{\mathbf{x}}^i - \mathbf{x}^i).$$

Bei der Verwendung von h"oheren Ansatzr"aumen wird der Vanka-Gl"atter, insbesondere bei dreidimensionalen Problemen, schnell teuer.

Werden Finite Elemente gleicher Ordnung f"ur Druck und Geschwindigkeit verwendet (z.B. das $Q^1 - Q^1$ Element) ist eine alternative Blockung durch Umsortierung der Freiheitsgrade m"oglich. Wir schreiben den L"osungsvektor als

$$\mathbf{x}_h := \begin{pmatrix} \mathbf{x}^1 \\ \mathbf{x}^2 \\ \vdots \\ \mathbf{x}^N \end{pmatrix}, \quad \text{mit } \mathbf{x}^i := \begin{pmatrix} \mathbf{v}_x^i \\ \mathbf{v}_y^i \\ \mathbf{p}^i \end{pmatrix}, \quad i = 1, \ldots, N,$$

fassen also stets alle Freiheitsgrade in einem Knoten zusammen. Die Matrix wird entsprechend eingeschr"ankt:

$$\mathbf{X}_h = (\mathbf{X}_h^{ij})_{i,j=1}^N, \quad \mathbf{X}_h^{ij} = \begin{bmatrix} \mathbf{A}_{xx}^{ij} & \mathbf{A}_{xy}^{ij} & \mathbf{B}_x^{ij} \\ \mathbf{A}_{yx}^{ij} & \mathbf{A}_{yy}^{ij} & \mathbf{B}_y^{ij} \\ -\mathbf{B}_x^{ij} & -\mathbf{B}_y^{ij} & 0 \end{bmatrix}.$$

Auf die Matrix $\mathbf{X}_h$ kann nun zum Beispiel eine ged"ampfte Gau"s-Seidel oder eine ILU-Iteration

$$\mathbf{x}_h^{t+1} = \mathbf{x}_h^t + \theta \mathcal{ILU}(\mathbf{X}_h)^{-1}(\mathbf{b}_h - \mathbf{X}_h \mathbf{x}_h^{(t)})$$

angewendet werden, wobei $\mathcal{ILU}(\mathbf{X}_h)$ eine unvollst"andige LU-Zerlegung der Matrix $\mathbf{X}_h$ ist. Dabei behandeln wir die einzelnen Bl"ocke $\mathbf{X}_h^{ij}$ exakt. Durch diese Blockung wird lokal dem Charakter des Sattelpunktproblems Rechnung getragen. Der Gl"atter ist sehr effizient, kann insbesondere einfach auf Probleme erweitert werden, die neben den Navier-Stokes Gleichungen noch weitere Gleichungen beinhalten. Diese Art der Blockung setzt allerdings einen *equal-order Ansatz* f"ur Druck und Geschwindigkeit voraus.

Mehrgitter als Vorkonditionierer Das Design eines robusten Mehrgitter-Verfahrens f"ur die Navier-Stokes Gleichungen ist schwierig, insbesondere wenn die Reynoldszahl gr"o"ser wird, oder wenn anisotrop gestreckte Elemente (d.h. Elemente, bei denen das Verh"altnis zwischen Umkreis- und Innkreisradius asymptotisch nicht beschr"ankt ist) verwendet werden, was zum Aufl"osung von Grenzschichten zwingend erforderlich ist. Daher wird das Mehrgitterverfahren meist nicht direkt als L"oser, sondern als Vorkonditionierer in einem Krylow-Raum-Verfahren, z.B. GMRES verwendet. Bezeichnen wir mit $\mathcal{M}$ die Mehrgitter-Iteration, also mit

$$\mathbf{x}^{t+1} = \mathbf{x}^t + \mathcal{M}(\mathbf{b} - \mathbf{X}\mathbf{x}^t)$$

einen Mehrgitterschritt, so muss f"ur das Spektrum

$$\sigma(I - \mathcal{M}\mathbf{X}) < 1,$$

gelten, damit das Mehrgitterverfahren als L"oser robust konvergiert. Bei der Verwendung als Vorkonditionierer ist lediglich eine obere Schranke f"ur die Kondition:

$$\mathrm{cond}_2(\mathcal{M}\mathbf{X}) \leqslant C$$

erforderlich. Das vorkonditionierte System

$$\mathcal{M}\mathbf{X}\mathbf{x} = \mathcal{M}\mathbf{b}$$

kann dann z.B. mit dem GMRES-Verfahren gel"ost werden. Jede Matrix-Vektor-Multiplikation mit $\mathcal{M}\mathbf{X}$ erfordert dann die Anwendung des Mehrgitteralgorithmus.

5.3 L”osung der instation”aren Navier-Stokes Gleichungen

Wir betrachten die algebraischen Probleme, welche durch Zeitdiskretisierung der Navier-Stokes Gleichungen mit einem Zeitschrittverfahren (z.B. Crank-Nicolson oder Teilschritt-Theta-Verfahren) entstehen. Analog zu (5.14) schreiben wir:

$$\begin{bmatrix} \mathbf{A}(\tilde{\mathbf{v}}) & \mathbf{B} \\ -\mathbf{B}^T & 0 \end{bmatrix} \begin{bmatrix} \mathbf{v} \\ \mathbf{p} \end{bmatrix} = \begin{bmatrix} \mathbf{b} \\ \mathbf{g} \end{bmatrix}, \tag{5.16}$$

wobei $\mathbf{b}$ und $\mathbf{g}$ s”amtliche expliziten Anteile der Zeitdiskretisierung enthalten. F”ur die Matrix $\mathbf{A}(\tilde{\mathbf{v}})$ gilt nun bei Newton-Linearisierung:

$$\mathbf{A}(\tilde{\mathbf{v}}) = \Big((\phi^j, \phi^i) + k\theta\nu(\nabla\phi^j, \nabla\phi^i) + k\theta(\tilde{\mathbf{v}}_h \cdot \nabla\phi_h^j, \phi_h^i) + k\theta(\phi_h^j \cdot \nabla\tilde{\mathbf{v}}_h, \phi_h^i) \Big)_{i,j=1}^{N_v},$$

oder in kompakter Schreibweise:

$$\mathbf{A}(\tilde{\mathbf{v}}) = \mathbf{M} + k\theta\nu\mathbf{L} + k\theta\mathbf{C}(\tilde{\mathbf{v}}),$$

wobei $\mathbf{M}$ die Geschwindigkeits-Massematrix, $\mathbf{L}$ die Laplace- und $\mathbf{C}$ die Konvektionsmatrix ist. F”ur sehr kleine Zeitschritte $k \to 0$ entspricht jeder Zeitschritt von (5.16) einer L^2-Projektion in den Raum der schwach divergenzfreien Funktionen:

$$\begin{bmatrix} \mathbf{M} & \mathbf{B} \\ -\mathbf{B}^T & 0 \end{bmatrix} \begin{bmatrix} \mathbf{v} \\ \mathbf{p} \end{bmatrix} = \begin{bmatrix} \mathbf{b} \\ \mathbf{g} \end{bmatrix}.$$

Von einem effizienten L”osungsverfahren wird erwartet, dass dieses f”ur kleine Zeitschritte die spezielle Struktur optimal ausnutzt. Wir betrachten wieder das Schur-Komplement von (5.16)

$$\mathbf{B}^T\mathbf{A}(\tilde{\mathbf{v}})^{-1}\mathbf{B}\mathbf{p} = \mathbf{g} + \mathbf{B}^T\mathbf{A}(\tilde{\mathbf{v}})^{-1}\mathbf{b}.$$

Zur L”osung wollen wir eine *vorkonditionierte Richardson-Iteration* verwenden:

$$\begin{aligned} \mathbf{p}^n &= \mathbf{p}^{n-1} + \omega\mathcal{P}^{-1}(\mathbf{g} + \mathbf{B}^T\mathbf{A}(\tilde{\mathbf{v}})^{-1}\mathbf{b} - \mathbf{B}^T\mathbf{A}(\tilde{\mathbf{v}})^{-1}\mathbf{B}\mathbf{p}^{n-1}) \\ &= \mathbf{p}^{n-1} + \omega\mathcal{P}^{-1}\big(\mathbf{g} + \mathbf{B}^T\mathbf{A}(\tilde{\mathbf{v}})^{-1}(\mathbf{b} - \mathbf{B}\mathbf{p}^{n-1})\big). \end{aligned}$$

mit einem D”ampfungsfaktor ω und einem Vorkonditionierer $\mathcal{P}$ welchen wir sp”ater diskutieren. Wir schreiben die Iteration in einem zweistufigen Verfahren:

Algorithmus 5.10 (Richardson-Iteration zur L”osung der instation”aren Navier-Stokes Gleichungen)**.** Sei $\mathbf{p}^0 = 0$, $\omega > 0$ und $\mathcal{P}$ gegeben. Iteriere f”ur $n \geqslant 1$:

1. Geschwindigkeits-Schritt
$$\mathbf{A}(\mathbf{v}^{n-1})\mathbf{v}^n = \mathbf{b} - \mathbf{B}\mathbf{p}^{n-1}$$

2. Druck-Schritt
$$\mathcal{P}(\mathbf{p}^n - \mathbf{p}^{n-1}) = \omega(\mathbf{g} + \mathbf{B}^T\mathbf{v}^n)$$

Die Effizienz des Verfahrens h"angt von der ad"aquaten L"osung beider Teilschritte ab. Der Charakter der Gleichungen "andert sich stark mit der gew"ahlten Zeitschrittweite. Dies muss auch bei der L"osung ber"ucksichtigt werden.

5.3.1 Der Geschwindigkeits-Schritt

F"ur gro"se Zeitschrittweiten und moderate Reynoldszahlen ist die Gleichung

$$\mathbf{A}(\mathbf{v}^{n-1})\mathbf{v}^{n} = \mathbf{b},$$

eine Diffusions-Transportgleichung mit zwei oder drei (je nach Raumdimension) L"osungskomponenten. Diese Gleichung kann mit Mehrgitter-Verfahren effizient gel"ost werden. Insbesondere bei dominanter Konvektion ist die Wahl eines robusten Mehrgittergl"atters wesentlich. F"ur kleine Zeitschrittweiten "andert sich der Charakter der Matrix $A(\mathbf{v}^{n-1})$ und die Masse-Matrix wird dominant. Lediglich die Konvektionsmatrix $\mathbf{C}(\tilde{\mathbf{v}})$ ist nicht-diagonal in dem Sinne, dass hier die verschiedenen Geschwindigkeits-Komponenten miteinander koppeln. F"ur kleine Zeitschritte kann diese Matrix vereinfacht werden. Wir schreiben:

$$\mathbf{C}(\tilde{\mathbf{v}}) = \mathbf{C}_1(\tilde{\mathbf{v}}) + \mathbf{C}_2(\tilde{\mathbf{v}}) = \Big((\tilde{\mathbf{v}}_h \cdot \nabla\phi_h^j, \phi_h^i)\Big)_{i,j} + \Big((\phi_h^j \cdot \nabla\tilde{\mathbf{v}}_h, \phi_h^i)\Big)_{i,j}.$$

Der erste Anteil ist eine Block-Diagonal Matrix, zwischen den verschiedenen Geschwindigkeitskomponenten bestehen keine Kopplungen. Hier sind also nur die Diagonalbl"ocke besetzt. Die Matrix $\mathbf{C}_2$ ist voll besetzt, die drei Geschwindigkeitskomponenten koppeln miteinander. Wenn wir den Anteil $\mathbf{C}_2$ *explizit* behandeln, so vereinfacht sich das System und kann entkoppelt in drei Teilschritten gel"ost werden:

$$\begin{bmatrix} \tilde{\mathbf{A}}(\tilde{\mathbf{v}}^{n-1})_1 & 0 & 0 \\ 0 & \tilde{A}(\tilde{\mathbf{v}}^{n-1})_2 & 0 \\ 0 & 0 & \tilde{A}(\tilde{\mathbf{v}}^{n-1})_3 \end{bmatrix} \begin{bmatrix} \mathbf{v}_1^n \\ \mathbf{v}_2^n \\ \mathbf{v}_3^n \end{bmatrix} = \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \\ \mathbf{b}_3 \end{bmatrix} - \mathbf{C}_2(\tilde{\mathbf{v}})\mathbf{v}^{n-1},$$

mit

$$\tilde{\mathbf{A}}(\tilde{\mathbf{v}}) = \mathbf{M} + k\theta\nu\mathbf{L} + k\theta\mathbf{C}_1(\tilde{\mathbf{v}}).$$

Die drei Matrizen unterscheiden sich dabei nur in gegebenenfalls unterschiedlichen Dirichlet-Vorgaben. F"ur sehr kleine Zeitschrittweiten k"onnen die einzelnen Gleichungen mit dem CG-Verfahren gel"ost werden, da die Masse-Matrix gut konditioniert ist.

5.3.2 Der Druck-Schritt

Der Druck-Schritt wird von der richtigen Wahl des Vorkonditionierers $\mathcal{P}$ bestimmt. Diese Wahl ist auch entscheidend f"ur die Konvergenz des Gesamtverfahrens. Wir bilden $\mathcal{P}$ mit einem additiven Ansatz

$$\mathcal{P}^{-1} = \alpha_M \mathcal{P}_M^{-1} + \alpha_L \mathcal{P}_L^{-1} + \alpha_C \mathcal{P}_C^{-1},$$

der den einzelnen Anteilen Rechnung tr"agt. Dabei nennen wir $\mathcal{P}_M$ den reaktiven, $\mathcal{P}_L$ den diffusiven und $\mathcal{P}_C$ den konvektiven Vorkonditionierer. Weiter sind $\alpha_M, \alpha_L, \alpha_C \geqslant 0$ D"ampfungsparameter. Die Druck-Korrektur wird dann in drei Schritten durchgef"uhrt:

$$\begin{aligned}
\mathcal{P}_M \mathbf{p}_M^\delta &= \mathbf{g} + \mathbf{B}^T \mathbf{v}^n \\
\mathcal{P}_L \mathbf{p}_L^\delta &= \mathbf{g} + \mathbf{B}^T \mathbf{v}^n \\
\mathcal{P}_C \mathbf{p}_C^\delta &= \mathbf{g} + \mathbf{B}^T \mathbf{v}^n \\
\mathbf{p}^n &= \mathbf{p}^{n-1} + \omega(\mathbf{p}_M^\delta + \mathbf{p}_L^\delta + \mathbf{p}_C^\delta).
\end{aligned}$$

Der diffusive Vorkonditionierer $\mathcal{P}_L$: F"ur gro"se Zeitschrittweiten und moderate Reynoldszahlen ist die Laplace-Matrix dominant:

$$\mathbf{B}^T \mathbf{A}(\tilde{\mathbf{v}})^{-1} \mathbf{B} \sim \frac{1}{k\theta\nu} \mathbf{B}^T \mathbf{L}^{-1} \mathbf{B}.$$

Diese Matrix ist gerade das klassische Schur-Komplement $\mathbf{\Sigma}_h$ der Stokes-Gleichungen. Es gilt $\mathrm{cond}_2(\mathbf{M}_p^{-1} \mathbf{\Sigma}_h) \sim 1$ mit der Massematrix $\mathbf{M}_p$ des Druckraums. Diese Matrix ist also ein optimaler Vorkonditionierer f"ur den diffusiven Anteil:

$$\mathcal{P}_L := \frac{1}{k\theta\nu} \mathbf{M}_p.$$

Die Invertierung der Massematrix kann mit einem CG- oder Mehrgitterverfahren in optimaler Komplexit"at erfolgen.

Der reaktive Vorkonditionierer $\mathcal{P}_M$: F"ur sehr kleine Zeitschrittweiten ist die (Geschwindigkeits) Masse-Matrix im Schurkomplement dominant:

$$\mathbf{B}^T \mathbf{A}(\tilde{\mathbf{v}})^{-1} \mathbf{B} \sim \mathbf{B}^T \mathbf{M}_\nu^{-1} \mathbf{B}.$$

Eine heuristische Analyse zeigt wegen

$$\mathbf{B}^T \mathbf{M}_\nu^{-1} \mathbf{B} \sim \mathrm{div} \circ \mathrm{id}^{-1} \circ \mathrm{grad} \sim \Delta$$

eine N"ahe zum Laplace-Operator. Verwendet man f"ur den reaktiven Vorkonditionierer $\mathcal{P}_M$ allerdings die Laplace-Matrix im Druck-Ansatzraum Q_h, so werden fehlerhafte Druck-Randwerte (von Dirichlet- oder Neumann-Art) eingef"uhrt. Stattdessen wird der Vorkonditionierer explizit aufgebaut als

$$\mathcal{P}_M := \mathbf{B}^T \bar{\mathbf{M}}_\nu^{-1} \mathbf{B},$$

wobei $\bar{\mathbf{M}}_\nu$ die *gelumpte Masse-Matrix* im Geschwindigkeitsraum V_h ist. Diese Matrix wird mit der Trapez-Regel als Quadraturformel aufgestellt und ist eine Diagonalmatrix. Die Invertierung ist demnach trivial. Im Allgemeinen ist die Besetzungsstruktur von $\mathcal{P}_M$ ist gr"o"ser als die "ubliche Laplace-Matrix. Das Problem kann jedoch optimal mit einem Mehrgitter-Verfahren gel"ost werden. F"ur das rotierte $\tilde{Q}^1 - Q^0$ Element ist die Matrix $\mathcal{P}_M$ gerade der "ubliche 5-Punkte Stern, jedoch mit speziellen Randwerten.

Der konvektive Vorkonditionierer $\mathcal{P}_C$: F"ur konvektionsdominante Probleme existiert kein einfacher optimaler Vorkonditionierer:

$$\mathbf{B}^T\mathbf{A}(\tilde{\mathbf{v}})^{-1}\mathbf{B} \sim \frac{1}{k\theta}\mathbf{B}^T\mathbf{C}(\tilde{\mathbf{v}})^{-1}\mathbf{B}.$$

Eine M"oglichkeit besteht darin, die Inverse der Konvektionsmatrix durch eine unvollst"andige LU-Zerlegung zu ersetzen:

$$\mathcal{P}_C := \frac{1}{k\theta}\mathbf{B}^T\mathcal{ILU}(\mathbf{C}(\tilde{\mathbf{v}}))^{-1}\mathbf{B}$$

Das Problem k"onnte dann mit einem Mehrgitter oder GMRES-Verfahren gel"ost werden. Bei der L"osung des Geschwindigkeitsschritts kann die ILU der Matrix $\mathbf{A}(\tilde{\mathbf{v}})$ als robuster Gl"atter eingesetzt werden. Diese, bereits erstellte Zerlegung kann hier auch alternativ als Konvektions-Vorkonditionierer eingesetzt werden. Im Allgemeinen versucht man jedoch ohne Vorkonditionierung des Konvektionsterms auszukommen, also mit der "au"serst einfachen Wahl

$$\mathcal{P}_c = I.$$

6 Kompressible Str”omungen

Wir haben bisher die inkompressiblen Navier-Stokes bzw. Stokes-Gleichungen behandelt. Die Grundannahme der Inkompressiblit”at bedeutet, dass sich die Dichte auch bei gr”o”seren Kr”aften nicht ”andert. Inkompressiblit”at ist eine gute Approximation f”ur Wasser, aber zum Beispiel auch f”ur Luftstr”omungen bei kleinen Geschwindigkeiten.

Dichte”anderungen k”onnen verschiedene Ursachen haben, welche unterschieden werden m”ussen. Bei einer echt kompressiblen Str”omung geht man davon aus, dass das Volumen durch Kr”afte komprimiert oder expandiert wird. Dies ist z.B. der Fall bei Luftstr”omungen bei gro”sen Geschwindigkeiten, etwa bei der Umstr”omung von schnell fliegenden Flugzeugen. Dichte”anderungen k”onnen jedoch auch durch externe Faktoren wie Temperaturschwankungen, oder unterschiedliche Zusammensetzung des Mediums stattfinden: die Dichte von Meerwasser h”angt einerseits von der Temperatur, jedoch auch vom Salzgehalt ab. Diese Dichte”anderungen m”ussen von echt kompressiblen Verhalten unterschieden werden.

Oftmals sind beide Effekte jedoch eng miteinander verbunden, denn ein schnelles expandieren oder komprimieren von Gasen f”uhrt zu einer Energie”anderung, welche sich in Temperatur-Effekten widerspiegelt. Wir werden daher zun”achst die kompressiblen Gleichungen aus Kapitel **??** wieder aufgreifen und zus”atzlich termische Aspekte in der Modellierung ber”ucksichtigen. Ab jetzt werden wir nicht mehr von *Isothermie* des Mediums ausgehen.

6.1 Modelle kompressibler Str”omungen

6.1.1 Energieerhaltung

Wir gehen hier noch einmal im Detail auf die Gleichung der Energieerhaltung ein. Zun”achst zitieren wir aus der Einleitung die Erhaltungsgleichungen f”ur Masse und Impuls:

$$\partial_t \rho + \operatorname{div}(\rho \mathbf{v}) = 0, \quad \rho \partial_t \mathbf{v} + \rho \mathbf{v} \cdot \nabla \mathbf{v} = \rho \mathbf{f} + \operatorname{div} \boldsymbol{\sigma}. \tag{6.1}$$

Physikalische Grundannahme ist die Erhaltung von kinetischer Energie

$$E_{kin}(V) = \frac{1}{2} \int_V \rho |\mathbf{v}|^2 \, dx,$$

und innerer Energie

$$E_{in}(V) = \int_V \rho e \, dx,$$

in einem materiellen Volumen $V \subset \mathbb{R}^3$. Die "Anderung der Energie entspricht der Summe aus wirkender mechanischer Leistung

$$P(V) = \int_V \rho \mathbf{f} \cdot \mathbf{v} \, dx + \int_{\partial V} \mathbf{n} \cdot \boldsymbol{\sigma} \mathbf{v} \, do,$$

im Volumen V und auf dem Rand ∂V sowie der Energiezufuhr durch W"armequellen h sowie dem Energiefluss auf dem Rand $\mathbf{q}$:

$$Z(V) = \int_V \rho h \, dx - \int_{\partial V} \mathbf{n} \cdot \mathbf{q} \, do.$$

Dabei gilt f"ur die Oberfl"achenterme bei Verwendung von $\boldsymbol{\sigma} = \boldsymbol{\sigma}^T$

$$\int_{\partial V} \mathbf{n} \cdot \boldsymbol{\sigma} \mathbf{v} \, do = \int_V \operatorname{div}(\boldsymbol{\sigma} \mathbf{v}) \, dx, \quad \int_{\partial V} \mathbf{n} \cdot \mathbf{q} \, do = \int_V \operatorname{div} \mathbf{q} \, dx.$$

Es gilt dann

$$d_t \{ E_{in}(V) + E_{kin}(V) \} = P(V) + Z(V).$$

Das Reynoldsche Transport, Theorem **??** liefert bei hinreichender Regularit"at

$$\left\{ \partial_t(\rho e) + \operatorname{div}(\rho e \mathbf{v}) \right\} + \left\{ \partial_t \left(\frac{1}{2} \rho |\mathbf{v}|^2 \right) + \operatorname{div} \left(\frac{1}{2} \rho |\mathbf{v}|^2 \mathbf{v} \right) \right\} = \left\{ \rho h - \operatorname{div} \mathbf{q} \right\} + \left\{ \rho \mathbf{f} \cdot \mathbf{v} + \operatorname{div}(\boldsymbol{\sigma} \mathbf{v}) \right\}. \tag{6.2}$$

Diese Gleichung l"asst sich mit Hilfe der Gleichungen f"ur Impulserhaltung und Masseerhaltung wesentlich vereinfachen. Es gilt:

$$\begin{aligned} \partial_t \left(\frac{1}{2} \rho |\mathbf{v}|^2 \right) &= \frac{1}{2} \partial_t \rho |\mathbf{v}|^2 + \rho \partial_t \mathbf{v} \cdot \mathbf{v} \\ \operatorname{div} \left(\frac{1}{2} \rho |\mathbf{v}|^2 \mathbf{v} \right) &= \frac{1}{2} |\mathbf{v}|^2 \operatorname{div}(\rho \mathbf{v}) + \rho \mathbf{v} \cdot \nabla \mathbf{v} \cdot \mathbf{v}. \end{aligned}$$

Zusammen folgt mit (6.1) und $\operatorname{div}(\sigma \mathbf{v}) = \boldsymbol{\sigma} : \nabla \mathbf{v} + \operatorname{div}(\boldsymbol{\sigma}) \cdot \mathbf{v}$

$$\partial_t \left(\frac{1}{2} \rho |\mathbf{v}|^2 \right) + \operatorname{div} \left(\frac{1}{2} \rho |\mathbf{v}|^2 \mathbf{v} \right) = \frac{1}{2} |\mathbf{v}|^2 \left(\underbrace{\partial_t \rho + \operatorname{div}(\rho \mathbf{v})}_{=0} \right) + \left(\underbrace{\rho \partial_t \mathbf{v} + \rho \mathbf{v} \cdot \nabla \mathbf{v}}_{= \rho \mathbf{f} + \operatorname{div} \boldsymbol{\sigma}} \right) \cdot \mathbf{v}$$

Hiermit reduziert sich die Gleichung der Energieerhaltung zu

$$\partial_t(\rho e) + \operatorname{div}(\rho e \mathbf{v}) = \rho h - \operatorname{div} \mathbf{q} + \boldsymbol{\sigma} : \nabla \mathbf{v}.$$

Wir fassen zun"achst die Erhaltungsgleichungen f"ur Masse (**??**), Impuls (**??**) und Energie zusammen:

$$\begin{aligned} \partial_t \rho + \nabla \cdot (\rho \mathbf{v}) &= 0, \\ \partial_t(\rho \mathbf{v}) + \nabla \cdot (\rho \mathbf{v} \otimes \mathbf{v}) - \nabla \cdot \boldsymbol{\sigma} &= \rho \mathbf{f} \\ \partial_t(\rho e) + \nabla \cdot (\rho e \mathbf{v}) - \boldsymbol{\sigma} : \nabla \mathbf{v} + \nabla \cdot \mathbf{q} &= \rho h, \end{aligned} \tag{6.3}$$

mit der Dichte $\rho : \Omega \to \mathbb{R}$, dem Geschwindigkeitsfeld $\mathbf{v} : \Omega \to \mathbb{R}^d$, einer wirkenden Volumenkraft $\mathbf{f} : \Omega \to \mathbb{R}^d$, der Energiedichte $e : \Omega \to \mathbb{R}$ und einer W"armequelle (bzw. Senke) $h : \Omega \to \mathbb{R}$ sowie dem Energiefluss $\mathbf{q} : \Omega \to \mathbb{R}^d$. Dies ist die *konservative Formulierung* der kompressiblen Navier-Stokes Gleichungen in den *konservativen Variablen* Massedichte ρ, Impuls $\rho\mathbf{v}$ und Energiedichte ρe. Wir werden sp"ater eine *nicht-konservative* Form der Gleichungen in den primitiven Variablen Dichte ρ, Geschwindigkeit $\mathbf{v}$ und Temperatur ϑ herleiten.

Zun"achst muss dieses System von Gleichungen jedoch mit Hilfe geeigneter Materialgesetze geschlossen werden. Im Fall Newtonscher Str"omungen gilt f"ur den Spannungstensor der lineare Zusammenhang

$$\boldsymbol{\sigma} = -pI + \rho\nu(\nabla\mathbf{v} + \nabla\mathbf{v}^T) - \frac{2}{3}\rho\nu(\nabla\cdot\mathbf{v})I. \tag{6.4}$$

Dabei ist $p : \Omega \to \mathbb{R}$ der Druck, welcher nun - je nach Modell - eine unterschiedliche physikalische Rolle einnehmen wird. Die kompressiblen Navier-Stokes Gleichungen sind kein Sattelpunktproblem, der Druck ist somit kein mathematisches Konstrukt zur Realisierung der Inkompressiblit"at. Stattdessen kommen im physikalische Eigenschaften zu. Wir werden auf den Zusammenhang zwischen diesem hydrostatischem und den anderen Erhaltungsgr"o"sen sp"ater eingehen. Ebenso werden wir Modelle f"ur den inneren W"armefluss $\mathbf{q} : \Omega \to \mathbb{R}^d$ angeben m"ussen.

6.1.2 Thermodynamische Modellierung

Das System der kompressiblen Navier-Stokes Gleichungen (6.3) zusammen mit dem Tensor (6.4) beinhalten in drei Dimensionen die 9 Unbekannten ρ, $\mathbf{v}$, p, e und $\mathbf{q}$. Demgegen"uber stehen 5 Gleichungen. Diese L"ucke wird durch weitere Materialgesetze geschlossen. Zun"achst fassen wir einige Modellannahmen zusammen:

- *1. Hauptsatz der Thermodynamik*: Innere Energie geht nicht verloren. Der Zuwachs an Energie in einem System entspricht der Summe aus eingebrachter W"arme δQ und Arbeit δW

  $$de = \delta Q + \delta W.$$

  Eingebrachte Arbeit kann z.B. die Kompression eines Volumens sein $\delta W = -p \cdot dV$, jedoch auch dissipative Arbeit wie innere Reibung.

- *2. Hauptsatz der Thermodynamik:* W"arme kann nicht ohne weitere Zustands"anderungen von einem K"orper niederer Temperatur auf einen K"orper h"oherer Temperatur "ubertragen werden. Dieser Hauptsatz ist zun"achst einleuchtend, da er der allgemeinen Anschauung entspricht. Seine Auswirkung ist jedoch weniger einfach zu erkl"aren. Der 2. Hauptsatz der Thermodynamik kann nun so verstanden werden, dass diese Beziehung eine neue Gr"o"se, die *Entropie* $s : \Omega \to \mathbb{R}$ beschreibt. Diese Entropie ist in jedem *irreversiblen Prozess* ansteigend. Hieraus kann gefolgert werden, dass ein *Perpetuum Mobile* nicht m"oglich ist.

Remark 6.1 (Entropie)**.** Die Entropie hat als physikalische Gr"o"se die Einheit $J/K = kg\,m^2/(Ks^2)$. Hieraus folgt sofort, dass die Entropie eine expansive Gr"o"se ist. Werden zwei Volumina mit fester Entropie zusammengef"ugt, so addiert sich auch die Entropie. Die Entropie kann bestimmt werden, so hat $1\,kg$ Wasser bei $10°C$ die Entropie $151J/K$ bei $20°C$ die Entropie $297J/K$ und schlie"slich bei $30°C$ die Entropie $437J/K$. Werden je $1\,kg$ kaltes Wasser und warmes Wasser vermischt, so k"onnte sich spontan (ohne Verrichtung weiterer Arbeit) ein Gemisch mit $20°C$ bilden, da die Summe der Entropien $151 + 437 = 588$ kleiner ist als die resultierende Entropie $297 + 297 = 594$. Ein Umkehrung dieses Prozesses, also eine spontane Trennung von $30°C$ warmen Wasser ist nicht m"oglich, dass die Entropie nur steigen kann.

Eine einfache Form der einwirkenden Arbeit ist die Kompression des Volumens V, gegeben durch $\delta W = -p \cdot dV$. Bei fester Masse $m(V) = \rho V = \text{const}$ gilt f"ur die Volumen"anderung $dV = d(\rho^{-1})$ und also $\delta W = -p d(\rho^{-1})$. Es folgt f"ur die Energie

$$de = \delta Q - p d(\rho^{-1}). \tag{6.5}$$

Wir betrachten nun den Fall, dass keine Arbeit einwirkt, und das Volumen konstant bleibt, also $dV = 0$. Dem System wird W"arme hinzugef"ugt wird. Beobachtung zeigt, dass dies proportional zum Anstieg der Temperatur $\vartheta : \Omega \to \mathbb{R}$ ist

$$\delta Q = c_v d\vartheta,$$

mit der *spezifischen W"arme bei konstantem Volumen* $c_v > 0$. Dann gilt f"ur die Energie

$$de = \delta Q = c_v d\vartheta \quad \Rightarrow \quad e = c_v \vartheta + \text{“const”}. \tag{6.6}$$

Die innere Energie ist in diesem Fall proportional zur Temperatur und einer Konstante. Diese neue Zustandsgr"o"se muss ebenso mit den bereits vorhandenen Gr"o"sen in Verbindung gesetzt werden. F"ur ein sogenannte *ideales Gas* (ein Gas, das homogen ist, also "uberall die gleiche Dichte hat) gilt

$$p = \frac{n}{V} \vartheta R_m,$$

mit der *universellen Gaskonstante* $R_m \approx 8.314 J/(mol\,K)$. Dabei ist n die Stoffmenge (in mol) und V das Volumen. Mit $\rho = m(V)/V$

$$\rho = \frac{m(V)}{V} = \frac{nM}{V},$$

und der *Molaren Masse* M folgt

$$p = \frac{R_m}{M} \rho \vartheta =: R_s \rho \vartheta, \tag{6.7}$$

mit der *spezifischen Gaskonstante* R_s. Hier gilt z.B. $R_s \approx 450$ f"ur Wasserdampf und $R_s \approx 2\,000$ f"ur Helium oder $R_s \approx 190$ f"ur CO_2.

Wir gehen nun davon aus, dass bei festem Druck W"arme hinzugef"ugt wird. Dann folgt aus (6.7)

$$p d(\rho^{-1}) = R_s d\vartheta.$$

F"ur die geleistete W"arme $\delta Q = c_p d\vartheta$, mit der *spezifischen W"arme bei konstantem Druck* $c_p > 0$ folgt dann mit (6.5) und (6.6)

$$de = c_v d\vartheta = c_p d\vartheta - R_s d\vartheta. \tag{6.8}$$

Somit gilt

$$R_s = c_p - c_v =: c_v(\gamma - 1) > 0 \quad \Rightarrow \quad \gamma = c_p/c_v. \tag{6.9}$$

F"ur zweiatomige Gase (z.B. O_2) gilt $\gamma = 1.4$. Die spezifische W"arme c_p hat die Einheit $J/(kg\,K)$ und kann f"ur verschiedene Stoffe bestimmt werden. Es gilt etwa $c_p \approx 1$ f"ur Luft oder $c_p \approx 15$ f"ur Wasserstoff.

Wir betrachten jetzt eine Expansion des Volumens ohne "au"sere W"armeeinfl"usse, also $\delta Q = 0$. Dann gilt mit (6.5), (6.7), (6.8) und (6.9)

$$\delta Q = 0 = de + p d(\rho^{-1}) = c_v d\vartheta + R_s \rho\vartheta d(\rho^{-1}) = c_v d\vartheta + c_v(\gamma - 1)\rho\vartheta d(\rho^{-1}),$$

und schlie"slich bei Division durch $\rho^{\gamma-1}$ und Zusammenfassen der Ableitungen

$$0 = \frac{d\vartheta}{\rho^{\gamma-1}} + (\gamma - 1)\rho\frac{\vartheta d(\rho^{-1})}{\rho^{\gamma-1}} = d\left(\frac{\vartheta}{\rho^{\gamma-1}}\right).$$

Also ist bei Verwendung des Gasgesetzes

$$\frac{\vartheta}{\rho^{\gamma-1}} = \text{"const"}, \quad \frac{p}{\rho^{\gamma}} = \text{"const"}.$$

Hieraus folgern wir das *Materialgesetz*

$$p = \alpha\rho^{\gamma},$$

mit eine Temperatur-abh"angigen Konstante $\alpha > 0$ und $\gamma = c_p/c_v = 1.4$.

Es bleibt, denn inneren W"armefluss $\mathbf{q}$ zu beschreiben. Ein einfaches Modell schreibt einen W"armestrom vor, welcher proportional zum Gradienten der Temperatur ist.

$$\mathbf{q} = -\kappa\nabla\vartheta,$$

mit dem *W"armeleitkoeffizienten* $\kappa > 0$. Der W"armeleitkoeffizient hat die Einheit $J/(s\,m\,K)$ und es gilt z.B. f"ur Wasserdampf oder Luft $\kappa \approx 0.025$ f"ur Wasserstoff $\kappa \approx 0.5$, f"ur Wasser $\kappa \approx$, f"ur Eisen $\kappa \approx 80$ und f"ur Silber $\kappa \approx 400$.

Schlie"slich folgt aus (6.8) mit (6.7) bei Vernachl"assigung des konstanten Anteils der Energie

$$\rho e = \rho c_v \vartheta = c_p \rho\vartheta - \rho R_s \vartheta = c_p \rho\vartheta - p. \tag{6.10}$$

Weiter gilt

$$\begin{aligned}\boldsymbol{\sigma} : \nabla\mathbf{v} = \rho\nu\left((\nabla\mathbf{v} + \nabla\mathbf{v}^T) - \frac{2}{3}\operatorname{div}\mathbf{v} I\right) : \nabla\nu - p(\operatorname{div}\mathbf{v}) \\ = \frac{1}{2}\rho\nu(\nabla\mathbf{v} + \nabla\mathbf{v}^T)^2 - \frac{2}{3}\rho\nu(\operatorname{div}\mathbf{v})^2 - p\operatorname{div}\mathbf{v}.\end{aligned}$$

Hieraus folgt die Energieerhaltungsgleichung in konservativer Form

$$\partial_t(c_p\rho\vartheta - p) + \operatorname{div}((c_p\rho\vartheta - p)\mathbf{v}) - \operatorname{div}(\kappa\nabla\vartheta) + p\operatorname{div}\mathbf{v} \\ - \frac{2}{3}\rho\nu(\nabla\mathbf{v} + \nabla\mathbf{v}^T)^2 + \frac{2}{3}\rho\nu(\operatorname{div}\mathbf{v})^2 = \rho h \quad (6.11)$$

6.1.3 Primitive Formulierung der kompressiblen Navier-Stokes Gleichungen und Vereinfachungen

Zur Herleitung der Energieerhaltungsgleichung in den primitiven Variablen Dichte ρ, Druck p, Geschwindigkeit $\mathbf{v}$ und Temperatur ϑ k"onnen wir wiefolgt umformen. Es gilt:

$$\partial_t(c_p\rho\vartheta - p) = c_p\vartheta\partial_t\rho + c_p\rho\partial_t\vartheta - \partial_t p,$$

sowie

$$\operatorname{div}((c_p\rho\vartheta - p)\mathbf{v}) = c_p\rho\mathbf{v}\cdot\nabla\vartheta + c_p\vartheta\operatorname{div}(\rho\mathbf{v}) - \mathbf{v}\cdot\nabla p - p\operatorname{div}\mathbf{v}.$$

Also bei Verwenden der Masseerhaltung

$$\partial_t(c_p\rho\vartheta - p) + \operatorname{div}((c_p\rho\vartheta - p)\mathbf{v}) = c_p\rho(\partial_t\vartheta + \mathbf{v}\cdot\nabla\vartheta) - \mathbf{v}\cdot\nabla p - p(\operatorname{div}\mathbf{v}).$$

Mit diesen weiteren Beziehungen k"onnen wir die kompressiblen Navier-Stokes Gleichungen in den primitiven Variablen Dichte ρ, Geschwindigkeit $\mathbf{v}$, Temperatur ϑ und Druck p formulieren. Es gilt

$$\begin{aligned} \partial_t\rho + \operatorname{div}(\rho\mathbf{v}) &= 0, \\ \rho\big(\partial_t\mathbf{v} + \mathbf{v}\cdot\nabla\mathbf{v}\big) - \rho_f\nu_f\operatorname{div}(\nabla\mathbf{v} + \nabla\mathbf{v}^T) - \frac{2}{3}\rho_f\nu_f\operatorname{div}\mathbf{v}I + \nabla p &= \rho\mathbf{f} \\ c_p\rho\big(\partial_t\vartheta + \mathbf{v}\cdot\nabla\vartheta\big) - \operatorname{div}(\kappa\nabla\vartheta) - \mathbf{v}\cdot\nabla p - \partial_t p & \\ -\frac{1}{2}\rho\nu(\nabla\mathbf{v} + \nabla\mathbf{v}^T)^2 + \frac{2}{3}\rho\nu(\operatorname{div}\mathbf{v})^2 &= \rho h. \end{aligned} \quad (6.12)$$

Dieses System wird geschlossen durch eine Gasgesetz, also z.B. dem allgemeinen Gasgesetz

$$p = \rho\vartheta R_s.$$

Unter gewissen Umst"anden kann das System der kompressiblen Navier-Stokes Gleichungen stark vereinfacht werden. Zun"achst gibt es F"alle, in denen Temperatur"anderung durch mechanische Kr"afte vernachl"assigt werden kann. Dies ist zum Beispiel bei langsameren Luftstr"omungen der Fall. Dann ist der einzig wirkende innere W"armestrom durch diffusive Effekte gegeben und die Gleichung der Energieerhaltung vereinfacht sich zu

$$c_p\rho(\partial_t\vartheta + \mathbf{v}\cdot\nabla\vartheta) - \operatorname{div}(\kappa\nabla\vartheta) = \rho h,$$

einer Diffusions-Transport Gleichung. F"ur ein ruhendes Volumen folgt bei $\mathbf{v} = 0$ mit der "ublichen W"armeleitungsgleichung

$$c_p\rho\partial_t\vartheta - \operatorname{div}(\kappa\nabla\vartheta) = \rho h,$$

das typische Modellproblem einer parabolischen partiellen Differentialgleichung.

6.1.4 Die Machzahl und "Ahnlichkeitsl"osungen

Zur Beurteilung von inkompressiblen Str"omungen haben wir zun"achst mit Hilfe einer Referenzgeschwindigkeit $\mathbf{v}^*$ und einer Referenzl"ange L die *Reynoldszahl*

$$Re = \frac{L\mathbf{v}^*}{\nu} = \frac{\rho^* L\mathbf{v}^*}{\mu}$$

eingef"uhrt, welche das Verh"altnis zwischen Reibungs- zu Tr"agheitskr"aften beschreibt. Weiter bezeichnet die *Froudezahl*

$$Fr = \frac{(\mathbf{v}^*)^2}{L\mathbf{f}},$$

das Verh"altnis zwischen Schwerkraft (da "ublicherweise $\mathbf{f} = -ge_3$ die Schwerkraft ist) und Tr"agheitskr"aften. Bei inkompressiblen Str"omungen mit homogener Dichte $\rho \equiv \rho^*$ spielt die Schwerkraft keine Rolle, da sie nur auf Dichteunterschiede wirkt. Im Fall kompressibler Str"omungen mit wechselnder Dichte "andert sich dies wesentlich.

In inkompressiblen Str"omungen wirken sich lokale "Anderungen unmittelbar global auf das ganze Str"omungsgebiet aus. Dies ist notwendig, da keine lokalen Dichte"anderungen m"oglich sind. Im Fall kompressibler Str"omungen werden wir sehen, dass auch Dichte- und Druck"anderungen einer endlichen Geschwindigkeit unterworfen sind. Hierzu betrachten wir eine starke Vereinfachung der Erhaltungsgleichungen:

- Wir betrachten eine isotherme Str"omung, so dass Temperatureffekte keine Rolle spielen.
- Wir betrachten nur kleine "Anderungen von Referenzdruck $|p - p^*| \ll |p^*|$ und Referenzdichte $|\rho - \rho^*| \ll |\rho^*|$.
- Wir betrachten ein im wesentlichen ruhendes Medium mit $\mathbf{v} \approx 0$ und $\nabla\mathbf{v} \approx 0$.

Unter diesen Vereinfachungen k"onnen die Erhaltungsgleichungen approximiert werden als

$$\begin{aligned} \partial_t \rho + \rho^* \operatorname{div} \mathbf{v} &= 0, \\ \rho^0 \partial_t \mathbf{v} + \nabla p &= 0. \end{aligned} \tag{6.13}$$

Wir gehen davon aus, dass Druck und Dichte einem differenzierbaren Zusammenhang $p = p(\rho)$ gen"ugen. Im Fall isothermer Str"omungen wird z.B. das *barotrope Gasgesetz* $p = \alpha\rho^\gamma$ mit $\alpha > 0$ und $\gamma = 1.4$ im Fall zweiatomiger Gase verwendet. Dann gilt:

$$\partial_t p = \frac{dp}{d\rho}\partial_t \rho$$

mit einem $c^2 > 0$. Wir gehen nun davon aus, dass $|p - p^*|$ und $|\rho - \rho^*|$ so klein sind, dass $c^2 := \frac{dp}{d\rho} > 0$ als konstant angesehen werden kann. Wir erhalten aus (6.13)

$$\partial_{tt} p = c^2 \partial_{tt} \rho = c^2 \operatorname{div}(\partial_t \mathbf{v}) = c^2 \rho^* \operatorname{div}(\nabla p) = c^2 \Delta p,$$

also eine Differenzialgleichung f"ur den Druck

$$\partial_{tt} p - c^2 \Delta p = 0,$$

und bei Annahme sehr kleiner St"orungen f"ur $p \approx p^* = c^2 \rho^* \approx \rho$ auch eine Differenzialgleichung f"ur die Dichte

$$\partial_{tt} \rho - c^2 \Delta \rho = 0.$$

Druck und Dichte gen"ugen demnach in N"aherung einer Wellengleichung. St"orungen werden mit der sogenannten *Schallgeschwindigkeit* $c > 0$ verbreitet.

Die Schallgeschwindigkeit kann aus dem Gasgesetz $p = p(\rho)$ bestimmt werden, h"angt also wesentlich vom Material ab. F"ur Luft gilt - stark abh"angig von der Temperatur - etwa $c_{\text{Luft}} \approx 340 \mathrm{m/s}$, f"ur Wasser etwa $c_{\text{Wasser}} \approx 1\,500 \mathrm{m/s}$.

Aus (6.13) erhalten wir durch einfaches Umformen

$$\partial_t p + c^2 \rho^* \operatorname{div} \mathbf{v} = 0.$$

Die Schallgeschwindigkeit spiegelt also auch den Grad an Inkompressibilit"at wieder. Die oben angegebenen Zahlenwerte zeigen, dass in Flu"ssigkeiten bei "ublicherweise sehr gro"ser Schallgeschwindigkeit ein inkompressibles Verhalten physikalisch sinnvoll ist. Bei Gasstr"omungen hingegen kann die Str"omungsgeschwindigkeit des Materials leicht die Gr"o"senordnung der Schallgeschwindigkeit $c \approx |\mathbf{v}|$ erreichen.

Zum Einfluss der Schallgeschwindigkeit betrachten wir weiter eine starke Vereinfachung der Gleichungen.

- Zun"achst folgern wir aus (6.13), dass $\partial_t \mathbf{v} \approx \nabla p$ die Zeitableitung der Geschwindigkeit dem Gradienten einer skalaren Funktion entspricht. Sie ist somit rotationsfrei $\partial_t \nabla \times \mathbf{v} = 0$ und wir folgern weiter $\nabla \cdot \mathbf{v} = 0$. Dann gilt ("Ubung)

$$\mathbf{v} \cdot \nabla \mathbf{v} = \frac{1}{2} \nabla |\mathbf{v}|^2 - \mathbf{v} \times (\nabla \cdot \mathbf{v}) = \frac{1}{2} \nabla |\mathbf{v}|^2.$$

- Weiter vernachl"assigen wir viskose Effekte, also $\nu = 0$.
- Wir einen station"aren Zustand mit $\partial_t \mathbf{v} = 0$, $\partial_t \rho = 0$ und $\partial_t p = 0$.
- Die Str"omung sei durch ein Volumenpotential getrieben, also durch $f = -\nabla V$, mit einer skalaren Funktion V.

Unter diesen Annahmen erhalten wir die vereinfachte Zustandsgleichungen

$$\rho \operatorname{div} \mathbf{v} + \mathbf{v} \cdot \nabla \rho = 0,$$
$$\frac{1}{2} \nabla |\mathbf{v}|^2 + c^2 \rho^{-1} \nabla \rho + \nabla V = 0,$$

wobei wir wieder $dp/d\rho = c^2$ mit der (lokalen) Schallgeschwindigkeit $c = c(p)$. Durch Multiplikation der zweiten Gleichung mit $\mathbf{v}\cdot$ und Ausnutzen der ersten kann die Dichte aus der Gleichung eliminiert werden:

$$\frac{1}{2}\mathbf{v}\cdot\nabla|\mathbf{v}|^2 - c^2(\operatorname{div}\mathbf{v}) + \mathbf{v}\cdot\nabla V = 0.$$

Es gilt

$$\frac{1}{2}\mathbf{v}\cdot\nabla|\mathbf{v}|^2 = \mathbf{v}\cdot(\mathbf{v}\cdot\nabla\mathbf{v}),$$

also transformiert sich die obige *Gasdynamische Gleichung* zu

$$\mathbf{v}\cdot(\mathbf{v}\cdot\nabla\mathbf{v}) - c^2\operatorname{div}\mathbf{v} + \mathbf{v}\cdot\nabla V = 0.$$

Das rotationsfreie Geschwindigkeitsfeld erlaubt die Einf"uhrung eines Geschwindigkeitspotentials $\Phi : \Omega \to \mathbb{R}$, so dass

$$\mathbf{v} = \nabla\Phi.$$

F"ur dieses Potential gilt die skalare Gleichung

$$\sum_{i,j=1}^{d} \partial_i\Phi\partial_j\Phi\partial_{ij}\Phi - c^2\Delta\Phi + \nabla\Phi\cdot\nabla V = 0. \tag{6.14}$$

Im Fall $c \gg |\mathbf{v}| = |\nabla\Phi|$ folgt die reine Potentialgleichung f"ur die Geschwindigkeit

$$-\Delta\Phi = 0.$$

Der Fall $|\mathbf{v}| \ll c$ bedeutet, dass die Ausbreitungsgeschwindigkeit von Schallwellen nahezu unendlich gro"s ist. Diese Annahme entspricht gerade der Inkompressibilit"at der Fl"ussigkeit.

Wir betrachten nun den allgemeinen Fall der Gleichung f"ur das Potential (6.14), schr"anken und aber auf den Fall von zwei r"aumlichen Dimensionen ein, so dass gilt

$$\nabla\Phi = \mathbf{v} =: \begin{pmatrix} u \\ w \end{pmatrix}.$$

Dann schreibt sich (6.14) als

$$(u^2 - c^2)\Phi_{xx} + (w^2 - c^2)\Phi_{yy} + 2uw\Phi_{xy} + uV_x + wV_y = 0.$$

Frieren wir u und v r"aumlich ein, so schreibt sich diese Gleichung als Differentialgleichung zweiter Ordnung mit Hauptteil $L(\Phi) = \nabla\cdot(A\nabla\Phi)$ mit

$$A = \begin{pmatrix} u^2 - c^2 & uw \\ uw & w^2 - c^2 \end{pmatrix}.$$

Zur Analyse des Typs dieser Gleichung betrachten wir die Eigenwerte des Hauptteils. Es gilt:

$$\lambda_1(A) = c^2, \quad \lambda_2(A) = c^2 - u^2 - w^2.$$

Wir unterscheiden folgende F"alle:

1. Beide Eigenwerte sind positiv, dann ist die Gleichung elliptisch:

$$c^2 > u^2 + w^2,$$

also

$$\frac{\sqrt{u^2 + w^2}}{c} = \frac{|\mathbf{v}|}{c} < 1.$$

Die Str"omungsgeschwindigkeit ist kleiner als die Schallgeschwindigkeit. Diesen Quotienten nennen wir die *Machzahl*

$$Ma = \frac{|\mathbf{v}|}{c}.$$

Str"omungen mit Machzahl kleiner 1 nennt man *subsonische Str"omungen*, oder *Unterschallstr"omungen*.

2. Ein Eigenwert ist positiv, einer ist negativ. Dann ist die Gleichung hyperbolisch, also vom Typ einer Transportgleichung

$$u^2 + w^2 > c^2,$$

also

$$Ma = \frac{|\mathbf{v}|}{c} > 1,$$

und die Str"omung wird *supersonisch* oder *"Uberschallstr"omung* genannt.

3. Schlie"slich bleibt der Fall eines positiven Eigenwerts und ein Eigenwert Null, also

$$u^2 + w^2 = c^2 \quad \Rightarrow \quad Ma \approx 1.$$

Diese Str"omung im "Ubergangsbereich wird *transsonische Str"omung* genannt und die Differentialgleichung ist vom parabolischen Typ.

Bei dieser Analyse ist zu ber"ucksichtigen, dass die Schallgeschwindigkeit und somit auch die Machzahl von den lokalen Werten von Druck, Dichte und Geschwindigkeit abh"angt. Ein Str"omungsproblem kann also in unterschiedlichen Bereichen der Str"omung ein unterschiedliches Verhalten annehmen. Das unterschiedliche Verhalten der Gleichung hat Einfluss auf die anzunehmenden Randwerte einer numerischen Simulation. Je ob supersonisch oder subsonisch m"ussen unterschiedliche S"atze von Randvorgaben f"ur Dichte und Druck vorgegeben werden.

6.1.5 Die Euler-Gleichungen

Bei schnellen Gasstr"omungen, z.B. bei der schnellen Umstr"omung eines Flugzeugs in gro"sen H"ohen k"onnen sehr gro"se Dichte"anderungen auftreten. Diese k"onnen fast zu Unstetigkeiten, sogenannten *Schocks* f"uhren. Entspricht die Fluggeschwindigkeit $\mathbf{v}$ etwa der Schallgeschwindigkeit c, d.h. im Fall $Ma \approx 1$ bauen sich Schockwellen auf. Diese Wellen sind als das *"Durchbrechen der Schallmauer"* bekannt. In diesen Str"omungsbereichen spielen Temperatureffekte oft eine geringere Bedeutung. Ebenso kann die innere Reibung

gegen"uber Tr"agheitseffekten oft vernachl"assigt werden. Wir werden daher zur N"aherung $\nu = 0$ ansetzen.

Beides "andert sich wieder bei noch gr"o"seren Geschwindigkeiten. Wird etwa der Wiedereintritt von Raumfahrzeugen in die Erdatmosph"are untersucht, so treten enorme Temperaturen auf, die sogar zu Bildung von Plasma (etwa bei den hohen Geschwindigkeiten die bei der R"uckkehr von einer Marsmission auftreten) f"uhren k"onnen. Da hier die Atmosph"are zum Bremsen genutzt wird ist die innere Reibung nat"urlich wesentlich.

Bei Str"omungen im Bereich hoher Machzahlen spielen die Erhaltungsgr"o"sen Dichte ρ, Impuls $\rho\mathbf{v}$ und Energie ρe die entscheidende Rolle, daher bietet sich eine Modellierung in konservativer Form an. Bei Vernachl"assigung von Reibung ergibt sich aus (6.3) f"ur die Masse- und Impulserhaltung Approximation

$$\begin{aligned} \partial_t \rho + \nabla \cdot (\rho \mathbf{v}) &= 0, \\ \partial_t (\rho \mathbf{v}) + \nabla \cdot (\rho \mathbf{v} \otimes \mathbf{v}) + \nabla \cdot (p I) &= \rho \mathbf{f}. \end{aligned}$$

Wir nehmen nun eine Energiegleichung hinzu und betrachten jetzt die totale Energie, gegeben durch

$$\rho E = \rho e + \frac{1}{2}\rho |\mathbf{v}|^2.$$

Es gilt in Erhaltungsform bei Vernachl"assigung der Reibung $\nu = 0$ und innerem W"armefluss $\mathbf{q} = 0$

$$\partial_t (\rho E) + \operatorname{div}(\rho E \mathbf{v} + p\mathbf{v}) = \rho \mathbf{f} \cdot \mathbf{v} + \rho h.$$

Zum Schlie"sen dieses Systems m"ussen wir weiter auf den Druck p eingehen. Wir gehen von einem idealen Gas aus, so dass gilt

$$p = \rho \vartheta R_s,$$

mit $R_s = c_p - c_\nu > 0$ sowie $\gamma = c_p/c_\nu > 1$. F"ur die innere Energie e gilt $e = c_\nu \vartheta$. Auf diese Weise kann der Druck p durch die anderen, konservativen Variablen ersetzt werden

$$Re = \frac{\rho^* L \mathbf{v}}{\mu}, \quad Fr = \frac{(\mathbf{v}^*)^2}{L}, \quad Pr = \frac{c_p \mu}{\kappa}, \quad Ma = \frac{\mathbf{v}^*}{\sqrt{\kappa R \vartheta^*}}.$$

Das System der *Euler-Gleichungen* ist von erster Ordnung in den konservativen Variablen ρ, $\rho\mathbf{v}$ sowie ρE. Es l"asst sich bei Fehlen externer Einfl"usse, also $h = 0$ sowie $\mathbf{f} = 0$ in einer reinen *Erhaltungsform* schreiben als

$$\partial_t \begin{pmatrix} \rho \\ \rho \mathbf{v} \\ \rho E \end{pmatrix} + \nabla \cdot \begin{pmatrix} \rho \mathbf{v} \\ \rho \mathbf{v} \otimes \mathbf{v} + p I \\ \rho E \mathbf{v} + p \mathbf{v} \end{pmatrix} = 0,$$

kurz f"ur $U = (\rho, \rho\mathbf{v}, \rho E)$

$$\partial_t U + \nabla \cdot F(U) = 0,$$

mit der Flussfunktion $F(\cdot)$. Diese spezielle Form wird wesentlich f"ur das Design numerischer Verfahren sein, welche Erhaltungseigenschaften auch in diskretisierter Form wiedergeben sollen.

Die Euler-Gleichungen sind das wesentliche System partieller Differentialgleichungen in der Aerodynamik. Als System von Differentialgleichungen erster Ordnung verhalten sich die L"osungen der Euler-Gleichung anders als L"osungen zu viskosen Str"omungsproblemen. Insbesondere die Frage der Regularit"at ist anders zu beantworten. Bei hohen Mach-Zahlen kann die L"osungen Unstetigkeiten in der Dichte, sogenannte Schocks aufweisen. In Abbildung 6.1 werden Euler-Str"omungen um eine Tragfl"ache bei verschiedenen Konfigurationen gezeigt. Wesentlicher Parameter ist immer die *freestream Machnumber*, also die Machzahl der umgebenen Str"omungen. Diese wirkt als Einflussbedingung an "au"seren R"andern im Einstr"ombereich. Da die Str"omung zum "Ausweichen" der Tragfl"ache beschleunigen muss, kann es auch bei $\mathrm{Ma} < 1$ lokale im Str"omungsgebiet zu Machzahlen gr"o"ser eins kommen. Die beiden Abbildungen auf der linken Seite korrespondieren zu einer subsonischen Str"omung. Hier gilt $\mathrm{Ma} < 1$ im gesamten Str"omungsgebiet. Aufgrund der Form der Tragfl"ache (diese ist nicht symmetrisch) stellt sich auf der Oberseite ein geringerer Druck ein. Dieser f"uhrt zu einer Auftriebskraft und erm"oglicht somit ein Fliegen. Die Abbildungen auf der rechten Seite werden auch durch eine subsonische Einstr"omung getrieben. Hier ist der Anstellwinkel $\alpha = 5$, die Tragfl"ache wird also von links unten angestr"omt. Hier erreicht die Str"omungsgeschwindigkeit lokal Machzahlen gr"o"ser eins $\mathrm{Ma} > 1$. Die nicht-viskose Eulerstr"omung bildet hier einen Schock, also eine Unstetigkeit in Machzahl, Druck und Dichte aus.

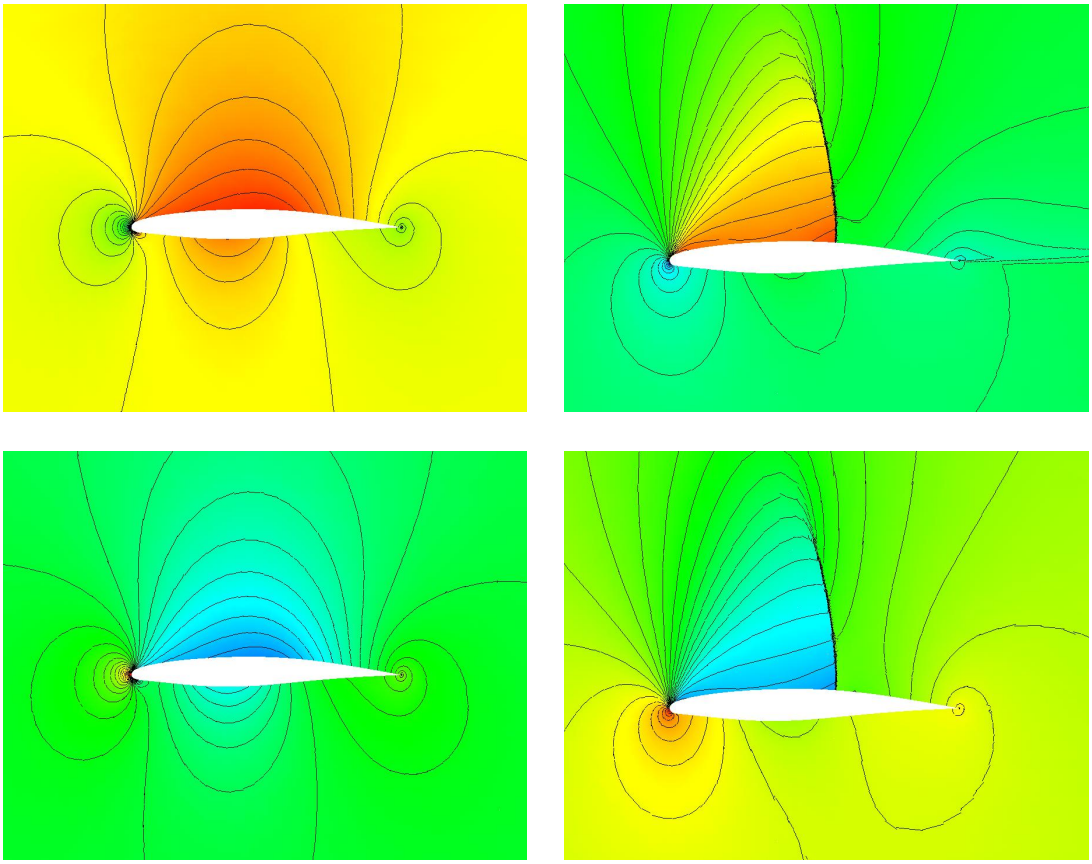


Figure 6.1: Eulerumstr"omung einer Tragfl"ache bei verschiedenen Anstr"omwinkeln α und verschiedenen Machzahlen im Fernfeld. Links: $\alpha = 0$ und $Ma = 0.6$. Rechts: $\alpha = 5$ und $Ma = 0.7$. Oben ist die lokale Machzahl dargestellt (dabei steht rot f"ur $Ma = 0.8$ auf der linken und $Ma = 1.6$ auf der rechten Seite, blau steht jeweils f"ur $Ma = 0$). Unten ist das Druckprofil angegeben (hier steht rot f"ur $p = 1.2$ links und 1.4 rechts sowie blau f"ur $p = 0.8$ links und $p = 0.2$ rechts).

6.1.6 Temperaturgetriebene Str"omungen

Neben den Euler-Gleichungen, denen die Annahme der Reibungsfreiheit zugrundeliegt, betrachten wir nun als weiteren Spezialfall thermisch getrieben Str"omungen: Dichte"anderungen werden im Wesentlichen durch Temperatur"anderungen, jedoch nicht durch kompressibles Verhalten der Str"omung selbst hervorgerufen. Dies bedeutet, dass die Str"omungsgeschwindigkeit gegen"uber der Schallgeschwindigkeit sehr klein ist, dass also $Ma \ll 1$ gilt. Hier k"onnen Temperatureffekte durch mechanische Wirkung vernachl"assigt werden. Wir gehen wieder von einem idealen Gasgesetz aus

$$p = \rho \vartheta R_s.$$

Weiter nehmen wir nun an, dass die Dichte im Wesentlichen von der Temperatur abh"angt und setzen die Relation $\rho = \rho(p, \vartheta) = \frac{p}{\vartheta R_s}$ in die Gleichung der Masseerhaltung ein. Es gilt

$$\partial_t \rho + \nabla \cdot (\rho \mathbf{v}) = 0 \quad \Rightarrow \quad \nabla \cdot \mathbf{v} + p^{-1}\big(\partial_t p + \mathbf{v} \cdot \nabla p\big) + \vartheta^{-1}\big(\partial_t \vartheta + \mathbf{v} \cdot \nabla \vartheta\big) = 0. \tag{6.15}$$

Im Fall gro"ser Dr"ucke $|p| \gg 1$ und gro"ser Temperaturen $|\vartheta| \gg 1$ liegt nur eine kleine St"orung der Inkompressibilit"at vor.

Diesen Fall werden wir nun im Grenzwert n"aher betrachten. Hierzu vereinfachen wir das System der kompressiblen Gleichungen unter den folgenden Annahmen:

- Reibung spielt keine Rolle.
- Die Str"omung ist isotherm $\vartheta \equiv \vartheta^*$
- Es liegt ein barotropes Gas vor $p = \alpha \rho^\gamma$.

Dann vereinfacht sich das System zu

$$\begin{aligned} \partial_t \rho + \nabla \cdot (\rho \mathbf{v}) &= 0 \\ \rho \partial_t \mathbf{v} + \rho \mathbf{v} \cdot \nabla \mathbf{v} + \nabla p &= 0 \\ p &= \alpha \rho^\gamma, \end{aligned}$$

mit Anfangsdaten $\mathbf{v} = \mathbf{v}^0$ und $\rho = \rho^0$ zum Zeitpunkt $t = 0$. Mit der Schallgeschwindigkeit c gilt

$$c^2 = \frac{dp}{d\rho} = \alpha\gamma\rho^{\gamma-1} \approx \alpha\gamma(\rho^0)^{\gamma-1}. \tag{6.16}$$

Wir gehen nun davon aus, dass die Machzahl klein ist

$$Ma = \frac{|\mathbf{v}|}{c} \approx \frac{|\mathbf{v}^*|}{\sqrt{\alpha\gamma(\rho^0)^{\gamma-1}}} \ll 1.$$

Betrachtet man etwa die Luftstr"omung in einem Zimmer (z.B. durch eine hei"se Heizung hervorgerufen) so gilt $\mathbf{v} \approx 1 m/s$ und $c \approx 300 m/s$, also $Ma \approx 0.003$.

Mit diesen Materialannahmen werden wir nun die Gleichung der Masseerhaltung gem"a"s (6.15) modifizieren. Es gilt

$$\partial_t p = \alpha\gamma\rho^{\gamma-1}\partial_t \rho \quad \Rightarrow \quad \partial_t \rho = \rho^{1-\gamma}\gamma^{-1}\alpha^{-1}\partial_t p,$$

sowie

$$\mathbf{v} \cdot \nabla p = \alpha \rho^{\gamma-1} \gamma \mathbf{v} \cdot \nabla \rho \quad \Rightarrow \quad \mathbf{v} \cdot \nabla \rho = \alpha^{-1} \gamma^{-1} \rho^{1-\gamma} \mathbf{v} \cdot \nabla p.$$

Also folgt f"ur die Gleichung der Masseerhaltung mit (6.16)

$$\nabla \cdot \mathbf{v} + \rho^{-\gamma} \alpha^{-1} \gamma^{-1} \big(\partial_t p + \mathbf{v} \cdot \nabla p) = c^{-2} \rho^{-1} (\partial_t p + \mathbf{v} \cdot \nabla p) + \nabla \cdot \mathbf{v} = 0.$$

F"ur gro"se Schallgeschwindigkeiten (also kleine Machzahlen) liegt wieder nur eine kleine St"orung der Inkompressibilit"at vor.

Wir betrachten nun den Fall der Euler-Gleichungen f"ur gro"se Schallgeschwindigkeiten. Es gilt f"ur Druck, Dichte und Geschwindigkeit

$$\begin{aligned} \partial_t \rho + \mathbf{v} \cdot \nabla \rho + \rho \nabla \cdot \mathbf{v} &= 0, \\ \partial_t \mathbf{v} + \mathbf{v} \cdot \nabla \mathbf{v} + \rho^{-1} \nabla p &= 0, \\ \partial_t p + c^2 \rho \nabla \cdot \mathbf{v} + \mathbf{v} \cdot \nabla p &= 0. \end{aligned} \tag{6.17}$$

Wir definieren (in einer Raumdimension) die Matrix

$$A(u) = A(\rho, \mathbf{v}, p) = \begin{pmatrix} \mathbf{v} & \rho & 0 \\ 0 & \mathbf{v} & \rho^{-1} \\ 0 & c^2 \rho & \mathbf{v} \end{pmatrix},$$

und schreiben das System kurz als

$$\partial_t u + A(u) \partial_x u = 0.$$

Diese Matrix hat die Eigenwerte

$$\lambda_1 = \mathbf{v}, \quad \lambda_{2/3} = \mathbf{v} \pm c.$$

Im Fall gro"ser Schallgeschwindigkeiten $c \gg 1$ ist diese Matrix sehr schlecht konditioniert. Im Fall einer langsamen Luftstr"omung mit $\rho = 1$, $\mathbf{v} = 1$ und $c = 300$ gilt $\mathrm{cond}(A) \approx 10^5$. Das System kann nicht mehr stabil gel"ost werden. Wir werden nun im Folgenden die Auswirkung von dieser schlechten Konditionierung n"aher untersuchen.

Dazu sei $\mathbf{v}^*$ eine Referenzgeschwindigkeit und L^* eine Referenzl"ange sowie ρ^* der Referenzdruck. Dann folgt mit $x^* = x/L^*$ und $t^* = \mathbf{v}^* t / L$ aus (6.17)

$$\begin{aligned} \partial_{t^*} \rho + \nabla^* \cdot (\rho \mathbf{v}) &= 0, \\ \rho \partial_{t^*} \mathbf{v} + \rho \mathbf{v} \cdot \nabla \mathbf{v} + \underbrace{(\rho^*)^{\gamma-1} (\mathbf{v}^*)^{-2}}_{=:\lambda^{-2}} \nabla p &= 0. \end{aligned} \tag{6.18}$$

F"ur den Vorfaktor des Drucks gilt

$$\lambda := \frac{(\mathbf{v}^*)^2}{(\rho^*)^{\gamma-1}} = \frac{\alpha\gamma(\mathbf{v}^*)^2}{\alpha\gamma(\rho^*)^{\gamma-1}} = \frac{\alpha\gamma(\mathbf{v}^*)^2}{c^2} = \alpha\gamma M a^2.$$

F"ur kleine Machzahlen wird der Vorfaktor immer gr"o"ser und die Bestimmung des Drucks aus dem Gasgesetz $p = \alpha\rho^\gamma$ ist nicht stabil, da hier unmittelbar eine Kopplung an die Dichte erfolgt.

F"ur kleine λ gehen wir nun von einer formalen Entwicklung des Drucks in Potenzen von λ aus

$$p = p_0 + \lambda p_1 + \lambda^2 p_2 + O(\lambda^3).$$

Einsetzen in (6.18) und Sortieren nach λ ergibt

$$\rho\partial_t \mathbf{v} + \rho\mathbf{v}\cdot\nabla\mathbf{v} + \nabla p_2 + \lambda^{-1}\nabla p_1 + \lambda^{-2}\nabla p_0 = O(\lambda).$$

Wir postulieren nun, dass diese Entwicklung im Grenz"ubergang $\lambda \to 0$ konvergiert. Dann erhalten wir zun"achst

$$\nabla p_0 = \nabla p_1 = 0,$$

sowie

$$\rho\partial_t \mathbf{v} + \rho\mathbf{v}\cdot\nabla\mathbf{v} + \nabla p_2 = 0.$$

Der Druck besteht aus zwei Anteilen p_0 und p_1 welche "ortlich konstant sind, jedoch von der Zeit abh"angen k"onnen. Wir nennen diese Teile den *thermodynamischen Druck*

$$p_{th}(t) = p_0(t) + \lambda^{-1} p_1(t).$$

Zur stabilen Approximation der Gleichungen bei kleinen Machzahlen wird nun die *Low-Mach-Number Approximation* verwendet. Der Druck wird aufgespalten in einen thermodynamischen Anteil p_{th}, welcher einzig in das Gasgesetz zur Bestimmung der Dichte eingeht, sowie einen "ortlich verteilten Anteil hydrodynamischen Anteil $p_{hyd}(x)$.

Die Low-Mach-Number Approximation Wir verwenden diese Approximationen nun im Fall kleiner Machzahlen zur Vereinfachung der kompressiblen Gleichungen. Der Druck sei aufgespalten in thermodynamische und hydrodynamische Anteile gem"a"s

$$p(x,t) = p_{th}(t) + p_{hyd}(x,t).$$

Dabei gehen wir davon aus, dass $|p_{hd}| \ll |p_{hd}|$. Es gelte das Gasgesetz in der Form

$$\rho = \frac{p_{th}}{R\vartheta}.$$

Neben der Aufteilung des Drucks gehen wir weiter davon aus, dass es nicht zu W"armeerzeugung durch mechanische Kr"afte kommt. Dann hat die Low-Mach-Number Approximation der "kompressiblen" Navier-Stokes die Form

$$\begin{aligned} \nabla\cdot\mathbf{v} - \frac{1}{\vartheta}\partial_t\vartheta - \frac{1}{\vartheta}\mathbf{v}\cdot\nabla\vartheta &= -\frac{1}{p_{th}}\partial_t p_{th}, \\ \rho(\partial_t\mathbf{v} + \mathbf{v}\cdot\nabla\mathbf{v}) - \operatorname{div}\left(\rho\nu(\nabla\mathbf{v} + \nabla\mathbf{v}^T) - \frac{2}{3}\rho\nu(\operatorname{div}\mathbf{v})I\right) + \nabla p_{hyd} &= \rho g, \\ \rho c_p(\partial_t\vartheta + \mathbf{v}\cdot\nabla\vartheta) - \kappa\Delta\vartheta &= \partial_t p_{th} + \rho h. \end{aligned} \tag{6.19}$$

Hier steht der thermodynamische Anteil des Drucks absichtlich auf der rechten Seite der Gleichungen. Diese skalarwertige Funktion kann mit Hilfe einer Anfangswertaufgabe gewonnen werden. Siehe z.B. [30].

6.2 Unstetige Galerkin-Verfahren

Wir betrachten die Euler-Gleichungen in Erhaltungsform

$$\begin{aligned} \partial_t \rho + \nabla \cdot (\rho \mathbf{v}) &= 0, \\ \partial_t (\rho \mathbf{v}) + \nabla \cdot (\rho \mathbf{v} \otimes \mathbf{v} + p I) &= \rho \mathbf{f}, \\ \partial_t (\rho E) + \nabla \cdot (\rho E \mathbf{v} + p \mathbf{v}) &= \rho h. \end{aligned}$$

Es handelt sich um eine nichtlineare Differentialgleichung erster Ordnung. Wir haben bereits gesehen, dass die Euler-Gleichungen auch bei glatten Daten unstetige L"osungen besitzen kann. Die "ubliche Finite Elemente Methode kann hier keine zufriedenstellende Ergebnisse liefern. Vereinfacht handelt es sich um den Extremfall eines reinen Transportproblems, siehe hierzu Abschnitt **??**. Der wesentliche Unterschied zur Stabilisierung von transportdominanten Str"omungen ist jedoch, dass dort nach wie vor Glattheit vorlag, jedoch mit gro"sen Gradienten. Hier ist auch die kontinuierliche L"osung nicht glatt.

6.2.1 Unstetige Galerkin-Verfahren f"ur die Laplace-Gleichung

Die Idee der unstetigen Galerkin-Verfahren ist die Konstruktion von Finite Elemente R"aumen, die keine globale Stetigkeit aufweisen. Stetigkeit der L"osung ist nicht stark in den Ansatzraum eingebaut sondern muss variationell erzwungen werden. Wir stellen diese Methode zun"achst zur Diskretisierung der Laplace-Gleichung vor. Auf einer Triangulierung Ω_h vom Gebiet Ω definieren wir den Raum

$$V_h := V_h^{(r),dg} := \{\phi \in L^2(\Omega),\ \phi\big|_K \in P^{(r)}(K)\},$$

wobei $P^{(r)}$ der Raum der Polynome von Grad $r \geqslant 0$ ist (entsprechend der Raum $Q^{(r)}$). Wir lassen hier explizit den Fall $r = 0$ zu. Die Basis von $V_h^{(r),dg}$ ist nicht zwangsl"aufig eine Knotenbasis. Da keine globale Stetigkeit erforderlich ist, kann auf jedem Element auch z.B. die Monombasis gew"ahlt werden

$$P_K^{(r)} := \operatorname{span}\left\{x^i y^j,\ 0 \leqslant i+j \leqslant r,\ 0 \leqslant i,j \leqslant r\right\}.$$

Aus Gr"unden der numerischen Stabilit"at (Konditionierung der Gram'schen Matrix) wird jedoch im Fall hoher Ansatzgrade eine andere, z.B. orthogonale Basis verwendet.

Eine einfache Galerkin-Diskretisierung der variationellen Formulierung

$$u \in H_0^1(\Omega) \quad (\nabla u, \nabla \phi) = (f, \phi) \quad \forall \phi \in H_0^1(\Omega),$$

im Raum $V_h^{(r),dg}$ schl"agt fehl. Da $V_h^{(r),dg} \not\subset H_0^1(\Omega)$ "ubertragen sich die Eigenschaften der Bilinearform - insbesondere die Elliptizit"at - nicht unmittelbar auf die Diskretisierung. Denn es gilt z.B. f"ur die Funktion

$$u_h^K = \begin{cases} 1 & x \in K, \\ 0 & x \notin K \end{cases}$$

$$(\nabla u_h^K, \nabla u_h^K) = (\nabla u_h^K, \nabla u_h^K)_K = 0.$$

Ebenso ist eine m"ogliche diskrete L"osung $u_h \in V_h$ von $(\nabla u_h, \nabla \phi_h)$ wegen

$$(\nabla u_h, \nabla \phi_h)_\Omega = \sum_{K \in \Omega_h} -(\Delta u_h, \phi_h)_K + \int_{\partial K} \mathbf{n} \cdot \nabla u_h \phi_h \, do = (f, \phi_h),$$

nicht unmittelbar eine gute Approximation der Laplace-Gleichung, da zus"atzliche Spr"unge "uber die Normale entstehen. Da sowohl $\mathbf{n} \cdot u_h$ als auch ϕ_h an den Zellkanten unstetig sind, wird im Limes $\mathbf{n} \cdot \nabla u_h = 0$ nicht erzwungen.

Wir starten daher mit einer neuen Herleitung einer variationellen Formulierung aufbauend auf der klassischen Form $-\Delta u = f$. Es gilt bei Multiplikation mit (unstetigen) Testfunktionen und Integration "uber das Gebiet

$$\sum_K (-\Delta u, \phi_h)_K = \sum_K (\nabla u, \nabla \phi_h)_K - \langle \mathbf{n} \cdot \nabla u, \phi_h \rangle_{\partial K} = (f, \phi_h) \quad \forall \phi_h \in V_h. \tag{6.20}$$

Neben den zellweisen Beitr"agen $(\nabla u, \nabla \phi_h)_K$ kommt ein "Fluss" der L"osung "uber die Kanten hinzu. Wollen wir die (als glatt angenommene) L"osung u durch eine diskrete und m"oglicherweise unstetige L"osung $u_h \in V_h$ ersetzen, so m"ussen wir uns auf jeder Kante ∂K entscheiden, ob u nun von links oder rechts betrachtet werden soll. Hierzu f"ugen wir die folgenden Notationen ein.

Definition 6.2 (Skelett eines Gitters)**.** Es sei Ω_h ein Finite Elemente Gitter. Unter dem *Skelett* des Gitters verstehen wir die Menge Γ_h aller Kanten e

$$\Gamma_h := \{e = \partial K_1 \cap \partial K_2, \ K_1, K_2 \in \Omega_h, \ K_1 \neq K_2\}.$$

Dabei unterscheiden wir zwischen innerem Skelett

$$\Gamma_{in} := \{e \in \Gamma_h, \ e \notin \partial\Omega\},$$

und "au"serem Skelett

$$\Gamma_a := \{e \in \Gamma_h, \ e \in \partial\Omega\}.$$

Jede Kante $e \in \Gamma_h$ hat einen Normalvektor $\mathbf{n}$ dessen Orientierung vom Betrachter abh"angt. Wird eine Kante e einer Zelle $K \in \Omega_h$ zugeordnet, also $e \in \partial K$, so ist mit $\mathbf{n}$ stets der nach au"sen gerichtete Normalvektor gemeint. F"ur eine (unstetige) Funktion $u : \Omega \to \mathbb{R}$

f"uhren wir - von K aus betrachtet - auf der Kante $e \in \partial K$ die Bezeichnungen u^+ sowie u^- ein

$$u^+(x) := \lim_{s \downarrow 0} u(x + s\mathbf{n}), \quad u^-(x) := \lim_{s \downarrow 0} u(x - s\mathbf{n}),$$

d.h., u^+ ist der Wert von u auf $e \in \partial K$ von au"sen gesehen und u^- der Wert von innen. Es ist also u^- die Spur von u auf ∂K von K aus gesehen. Weiter ist auf $e \in \partial K$ der Sprung einer Funktion gegeben durch

$$[u] := u^+ - u^-,$$

und der Mittelwert durch

$$\{\{u\}\} := \frac{1}{2}(u^+ + u^-).$$

Das Vorzeichen des Sprungs h"angt von der Orientierung der Normale ab. Zur Diskretisierung der variationellen Formulierung (6.20) ersetzen wir den Fluss $\mathbf{n} \cdot \nabla u$ durch eine *numerische Flussfunktion*, in welche beide Seiten von u_h eingehen k"onnen:

$$a_h(u_h, \phi_h) := \sum_{K \in \Omega_h} (\nabla u_h, \nabla \phi_h)_K + \int_{\partial K} h(u_h^+, u_h^-, \mathbf{n}) \phi_h \, do.$$

F"ur einen numerischen Fluss h definieren wir

Definition 6.3. Ein numerischer Fluss hei"st *konsistent*, falls f"ur die (glatte) L"osung $u = u^+ = u^-$ gilt

$$h(u, \mathbf{n})\phi = h(u^+, u^-, \mathbf{n})\phi = \mathbf{n} \cdot \nabla u \phi.$$

Hieraus folgern wir unmittelbar

Satz 6.4. Eine dG-Formulierung $a_h(u_h, \phi_h)$ ist konsistent, falls die numerische Flussfunktion h konsistent ist.

Ein nahe liegender Ansatz zur Definition des numerischen Flusses ist die Wahl des Mittelwerts, also

$$h(u_h^+, u_h^-, \mathbf{n}) = \mathbf{n} \cdot \{\{\nabla u_h\}\}.$$

Dieser Fluss ist nat"urlich konsistent. F"ur diese Flussfunktion konkretisieren wir die variationelle Formulierung. Da jede Kante doppelt auftaucht gilt mit dem st"uckweise definierten Gradienten

$$a_h(u_h, \phi_h) = (\nabla_h u_h, \nabla_h \phi_h) + \sum_{e \in \Gamma_h} \int_e \{\{\mathbf{n} \cdot \nabla u_h\}\} \cdot [\phi_h] \, do.$$

Zur k"urzeren Notation verwenden wir im Folgenden die Notation $\int_{|Gamma_h}$ f"ur die Summe aller Integrale "uber die Kanten:

$$a_h(u_h, \phi_h) = (\nabla_h u_h, \nabla_h \phi_h) + \int_{\Gamma_h} \{\{\mathbf{n} \cdot \nabla u_h\}\} \cdot [\phi_h] \, do.$$

Diese Formulierung ist scheinbar nicht symmetrisch! D.h., obwohl die variationelle Formulierung der Laplace-Gleichung ein Skalarprodukt definiert, gehen diese Eigenschaften in der diskreten Formulierung verloren. Symmetrie kann jedoch einfach durch Hinzunahme eines weiteren Flusses erzwungen werden. Wir definieren:

$$a_h(u_h, \phi_h) = (\nabla_h u_h, \nabla_h \phi_h) + \int_{\Gamma_h} \Big\{ \{\{\mathbf{n} \cdot \nabla u_h\}\} \cdot [\phi_h] + [u_h] \cdot \{\{\mathbf{n} \cdot \nabla \phi_h\}\} \Big\} \, do. \qquad (6.21)$$

F"ur die glatte L"osung u gilt $[u] = 0$, so dass auch diese Formulierung konsistent ist.

Eine entsprechend aufgestellte Systemmatrix A_h ist somit symmetrisch. Zur weiteren Analyse untersuchen wir die Regularit"at von A_h, also die Definitheit von $a_h(u_h, u_h)$. Es gilt

$$a_h(u_h, u_h) = \|\nabla_h u_h\|^2 + 2 \int_{\Gamma_h} \{\{\mathbf{n} \cdot \nabla u_h\}\}[u_h] \, do.$$

Wir w"ahlen nun eine Funktion $L^2(\Omega) \ni u_h^K \neq 0$ gem"a"s

$$u_h^K(x) := \begin{cases} 1 & x \in K, \\ 0 & x \notin K. \end{cases}$$

F"ur diese Funktion gilt $\nabla_h u_h^K = 0$ und somit $\mathbf{n} \cdot \nabla u_h = 0$, also schlie"slich $a_h(u_h, u_h) = 0$. Die Elliptizit"at vererbt sich bei unstetigen Galerkin-Verfahren nicht unmittelbar auf die Diskretisierung.

Zum Erreichen einer stabilen Diskretisierung reichern wir - "ahnlich den stabilisierten Finite Elemente Verfahren - die Flussfunktion weiter an. Wir definieren

$$a_h(u_h, \phi_h) = (\nabla_h u_h, \nabla_h \phi_h) + \int_{\Gamma_h} \Big\{ \delta [u_h][\phi_h] + \{\{\mathbf{n} \cdot \nabla u_h\}\}[\phi_h] + \{\{\mathbf{n} \cdot \nabla \phi_h\}\}[u_h] \Big\} \, do, \quad (6.22)$$

mit einem noch n"aher zu bestimmenden Parameter $\delta > 0$. Da f"ur die glatte L"osung $[u] = 0$ gilt ist auch diese Formulierung konsistent.

Satz 6.5. Die diskrete Bilinearform (6.22) ist stetig:

$$a_h(u, v) \leqslant c |\!|\!| u |\!|\!|_\delta |\!|\!| v |\!|\!|_\delta \quad \forall u, v \in H^2(\Omega_h) := \{\phi \in L^2(\Omega), \ \phi\Big|_K \in H^2(K)\},$$

mit einer Konstante $0 < c \leqslant 2$ und der dG-Norm

$$|\!|\!| v |\!|\!|_\delta := \sqrt{\|\nabla_h v\|^2 + \int_{\Gamma_h} \Big(\delta [v]^2 + \delta^{-1} \{\{\mathbf{n} \cdot \nabla v\}\}^2 \Big) \, do}.$$

Beweis: Es ist

$$a_h(u, v) = (\nabla_h u, \nabla_h v) + \int_{\Gamma_h} \Big(\delta [u][v] + \{\{\mathbf{n} \cdot \nabla u\}\}[v] + \{\{\mathbf{n} \cdot \nabla v\}\}[u] \Big) \, do.$$

Im Folgenden analysieren wir die einzelnen Anteile dieser Bilinearform.

(*i*) Zun"achst gilt

$$(\nabla_h u, \nabla_h v) \leqslant \sum_{K \in \Omega_h} \|\nabla u\|_K \, \|\nabla v\|_K \leqslant \left(\sum_K \|\nabla u\|_K^2 \right)^{\frac{1}{2}} \left(\sum_K \|\nabla v\|_K^2 \right)^{\frac{1}{2}} = \|\nabla_h u\| \, \|\nabla_h v\|.$$

(*ii*) F"ur die gemischten Terme gilt

$$\int_e \{\{\mathbf{n} \cdot \nabla u\}\}[v] \, do \leqslant \left(\int_e \delta^{-1} \{\{\mathbf{n} \cdot \nabla u\}\}^2 \, do \right)^{\frac{1}{2}} \left(\int_e \delta [v]^2 \, do \right)^{\frac{1}{2}},$$

also

$$\int_{\Gamma_h} \{\{\mathbf{n} \cdot \nabla u\}\}[v] \, do \leqslant \left(\int_{\Gamma_h} \delta^{-1} \{\{\mathbf{n} \cdot \nabla u\}\}^2 \, do \right)^{\frac{1}{2}} \left(\int_{\Gamma_h} \delta [v]^2 \, do \right)^{\frac{1}{2}},$$

und entsprechend f"ur den zweiten Term

$$\sum_{e \in \Gamma_h} \int_e \{\{\mathbf{n} \cdot \nabla v\}\}[u] \, do \leqslant \left(\int_{\Gamma_h} \delta^{-1} \{\{\mathbf{n} \cdot \nabla v\}\}^2 \, do \right)^{\frac{1}{2}} \left(\int_{\Gamma_h} \delta [u]^2 \, do \right)^{\frac{1}{2}}.$$

(*iii*) Schlie"slich folgt f"ur den verbleibenden Stabilisierungsterm

$$\int_{\Gamma_h} \delta [u][v]^2 \, do \leqslant \left(\int_{\Gamma_h} \delta [u]^2 \, do \right)^{\frac{1}{2}} \left(\int_{\Gamma_h} \delta [v]^2 \, do \right)^{\frac{1}{2}}.$$

(*iv*) Alle drei Bestandteile lassen sich durch das Produkt $|||u|||_\delta |||v|||_\delta$ beschr"anken, somit gilt mit einer Konstante $c \leqslant 2$

$$a_h(u, v) \leqslant c |||u|||_\delta |||v|||_\delta.$$

□

F"ur das weitere stellen wir zun"achst als Hilfsatz eine versch"arfte Spurabsch"atzung voraus:

Hilfsatz 6.6 (Spurabsch"atzung)**.** F"ur $u \in H^1(\Omega)$ gilt die Absch"atzung

$$\|u\|_{\partial K} \leqslant c \left(h_K^{-\frac{1}{2}} \|u\|_K + \|u\|_K^{\frac{1}{2}} \|\nabla u\|_K^{\frac{1}{2}} \right).$$

Beweis: "Ubung. □

Neben der Stetigkeit der Bilinearform zeigen wir nun noch die Elliptizit"at bzgl. der dG-Norm $|||\cdot|||_\delta$. Hier werden wir auch Bedingungen an den Stabilisierungsparameter $\delta > 0$ formulieren. Zur Herleitung der Elliptizit"at ben"otigen wir zun"achst einen Hilfsatz:

Hilfsatz 6.7. Es sei $e \in \partial K$ und $h := \operatorname{diam}(K)$. F"ur $v_h \in V_h^{(r),dg}$ gilt

$$\int_e \delta^{-1}\{\{\mathbf{n}\cdot\nabla v_h\}\}^2 \,do \leqslant c\frac{r^2}{\delta h_e}\|\nabla v_h\|_K^2,$$

wobei in $c > 0$ die Konstante der Spurabsch"atzung sowie die Konstante der inversen Absch"atzung eingeht.

Beweis: Es gilt mit Hilfe der modifizierten Spurabsch"atzung, Hilfsatz 6.6

$$\int_e \delta^{-1}\{\{\mathbf{n}\cdot\nabla v_h\}\}^2 \,do \leqslant \delta^{-1}\|\nabla v_h\|_e^2 \leqslant \frac{c}{\delta}\left(\frac{1}{h_k}\|\nabla v_h\|_K^2 + \|\nabla v_h\|\,\|\nabla^2 v_h\|\right).$$

F"ur die diskrete Funktion $v_h \in V_h^{(r),dg}$ verwenden wir die inverse Absch"atzung, Satz **??** und erhalten

$$\int_e \delta^{-1}\{\{\mathbf{n}\cdot\nabla v_h\}\}^2 \,do \leqslant \frac{c}{\delta}\left(\frac{1}{h_K} + \frac{cr^2}{h_K}\right)\|\nabla v_h\|_K^2.$$

□

F"ur die Wahl

$$\delta = \delta_0\frac{r^2}{h_K}$$

folgt

$$\int_e \delta^{-1}\{\{\mathbf{n}\cdot\nabla v_h\}\}^2 \,do \leqslant \frac{c^*}{\delta_0}\|\nabla v_h\|_K^2. \tag{6.23}$$

Hiermit k"onnen wir die Elliptizit"at der Bilinearform nachweisen:

Satz 6.8. F"ur $\delta_0 > 0$ gro"s genug existiert eine Konstante $\alpha > 0$ so dass gilt

$$a_h(u_h, u_h) \geqslant \alpha|||u_h|||_\delta^2.$$

Beweis: Wir betrachten f"ur $\alpha > 0$

$$\begin{aligned} a_h(u_h,u_h) - \alpha|||u_h|||_\delta^2 = (1-\alpha)\|\nabla_h u_h\|^2 + (1-\alpha)\int_{\Gamma_h}\delta[u_h]^2\,do \\ + 2\int_{\Gamma_h}\{\{\mathbf{n}\cdot\nabla u_h\}\}[u_h]\,do - \alpha\int_{\Gamma_h}\delta^{-1}\{\{\mathbf{n}\cdot\nabla u_h\}\}^2\,do. \end{aligned} \tag{6.24}$$

F"ur den gemischten Term gilt mit Cauchy Schwarz und Young'scher Ungleichung

$$2\int_{\Gamma_h}\{\{\mathbf{n}\cdot\nabla u_h\}\}[u_h]\,do \leqslant \epsilon\int_{\Gamma_h}\delta^{-1}\{\{\mathbf{n}\cdot\nabla u_h\}\}^2\,do + \frac{1}{\epsilon}\int_{\Gamma_h}\delta[u_h]^2\,do.$$

Zusammen mit (6.23) und (6.24) folgt

$$a_h(u_h,u_h) - \alpha|||u_h|||_\delta^2 = \left(1-\alpha-\frac{c^*}{\delta_0}(\alpha+\epsilon)\right)\|\nabla_h u_h\|^2 + \left(1-\alpha-\frac{1}{\epsilon}\right)\int_{\gamma_h}\frac{\delta_0 r^2}{h_K}[u_h]^2\,do$$

Zum Nachweis der Elliptizit"at m"ussen wir nun ein $\epsilon > 0$ sowie ein $\alpha > 0$ bestimmen, so dass gilt

$$1 - \alpha - \frac{c^*}{\delta_0}(\alpha + \epsilon) > 0, \quad 1 - \alpha - \frac{1}{\epsilon} > 0.$$

Die zweite Ungleichung fordert

$$0 < \alpha < 1 - \frac{1}{\epsilon},$$

also notwendigerweise $\epsilon > 1$ und $\alpha < 1$. Aus der ersten Ungleichung erhalten wir (mit $\epsilon > 1$)

$$0 < \alpha < \frac{1 - \frac{c^*}{\delta_0}\epsilon}{1 + \frac{c^*}{\delta_0}} \leqslant \frac{1 - \frac{c^*}{\delta_0}}{1 + \frac{c^*}{\delta_0}} = \frac{\delta_0 - c^*}{\delta_0 + c^*}.$$

D.h., f"ur z.B. $\delta_0 = 3c^*$ gilt f"ur jedes $\epsilon > 1$ Elliptizit"at mit jedem

$$\alpha < \frac{1}{2}.$$

□

Diese Methode wird die *Symmetric Interiour Penalty Galerkin* Formulierung, kurz SIPG genannt. In der Literatur findet sich eine Vielzahl von Modifikationen mit jeweils unterschiedlichen Eigenschaften. Wir beschr"anken uns hier jedoch exemplarisch auf das SIPG Verfahren.

Aus der Elliptizit"at und Symmetrie der Formulierung folgt unmittelbar, dass die Systemmatrix

$$\mathbf{A}_h = (a_{ij})_{i,j=1}^N, \; a_{ij} = a_h(\phi_h^j, \phi_h^i),$$

symmetrisch und positiv definit ist. D.h., die Matrix $\mathbf{A}_h$ ist insbesondere regul"ar und es existiert eine eindeutig bestimmte diskrete L"osung $u_h = \sum_i \mathbf{u}_i \phi_h^i$. Im nun Folgenden werden wir die Konvergenz dieser L"osung gegen die exakte L"osung $u \in H_0^1(\Omega)$ der Laplace-Gleichung untersuchen. Wir zeigen:

Satz 6.9 (SIPG)**.** Es sei $u \in H_0^1(\Omega) \cap H^{m+1}(\Omega)$ die L"osung der Laplace-Gleichung zu $f \in H^{m-1}(\Omega)$. Dann gilt f"ur $u_h \in V_h := V_h^{(r),dg}$ als L"osung von

$$a_h(u_h, \phi_h) = (f, \phi_h) \quad \forall \phi_h \in V_h,$$

wobei $a_h(\cdot,\cdot)$ durch (6.22) definiert ist die a priori Fehlerabsch"atzung

$$|||u - u_h|||_\delta \leqslant c h^{\min\{r,m\}} \|u\|_{H^{m+1}(\Omega)}.$$

Beweis: (*i*) Zun"achst sei $\mu := \min\{r, m\}$. Da die Elliptizit"at von $a_h(\cdot,\cdot)$ in Satz 6.8 f"ur diskrete Funktionen $\phi_h \in V_h$ bewiesen ist, f"ugen wir zun"achst einen Interpolationsfehler ein

$$u - u_h = \underbrace{u - i_h u}_{=:\eta} + \underbrace{i_h u - u_h}_{=:\xi_h}.$$

(*ii*) F"ur den Interpolationsfehler gilt bei $\delta = \delta_0/h$

$$|||\eta|||_\delta^2 = \sum_K \|\nabla_h(u - i_h u)\|_K^2 + \delta_0 \int_{\Gamma_h} h^{-1}[u - i_h u]^2 \, do + \delta_0^{-1} \int_{\Gamma_h} h\{\{\mathbf{n} \cdot \nabla(u - i_h u)\}\}^2 \, do.$$

Es ist bei Interpolation mit Finiten Elementen von Grad $r \geqslant 0$ unter Beachtung der jeweiligen Regularit"at

$$\|\nabla_h(u - i_h u)\| \leqslant c h^\mu \|\nabla^{\mu+1} u\|.$$

F"ur den Sprungterm gilt

$$h^{-1}\|[u - i_h u]\|_e^2 \leqslant c h^{-1} h^{2\mu+1} \|\nabla^{\mu+1} u\|_{P(e)}^2,$$

wobei $P(e)$ die beiden an e angrenzenden Elemente sind. So folgt

$$\delta_0 \int_{\Gamma_h} h^{-1}[u - i_h u]^2 \, do \leqslant c h^{2\mu} \|\nabla^{\mu+1} u\|^2.$$

Entsprechend gilt f"ur den Mittelwert

$$\delta_0^{-1} \int_{\Gamma_h} h\{\{\mathbf{n} \cdot \nabla(u - i_h u)\}\}^2 \, do \leqslant c h^{2\mu} \|\nabla^{\mu+1} u\|^2.$$

(*iii*) Wir sch"atzen nun den rein diskreten Anteil ξ_h ab. Mit Satz 6.8 gilt

$$\alpha |||\xi_h|||^2 \leqslant a_h(\xi_h, \xi_h) = a_h(u - u_h, \xi_h) - a_h(u - i_h u, \xi_h).$$

Aufgrund der Konsistenz der numerischen Flusses folgt $a_h(u - u_h, \xi_h) = 0$. Mit der Stetigkeit, Satz 6.5 folgt dann

$$|||\xi_h||| \leqslant c |||u - i_h u|||.$$

Kombination von dieser Absch"atzung mit den Interpolationsabsch"atzungen liefert das Ergebnis. □

Dieser Beweis zur a priori Absch"atzung kann mit den "ublichen Argumenten durchgef"uhrt werden, da trotz Nichtkonformit"at $V_h^{(r),dg} \not\subset H_0^1(\Omega)$ eine konsistente Formulierung vorliegt.

Die dG-Norm $|||\cdot|||_\delta$ entspricht im Wesentlichen der "ublichen Energienorm. Einzig der Term

$$\int_{\Gamma_h} \delta^{-1}\{\{\mathbf{n} \cdot \nabla u\}\}^2 \, do,$$

passt hinsichtlich der Regularit"at nicht in das "ubliche Konzept, da f"ur $u \in H^1(\Omega)$, die Spur der Normalableitung auf inneren Kanten nicht notwendigerweise existiert. Im Folgenden werden wir nun noch eine Fehlerabsch"atzung in der L^2-Norm herleiten.

Satz 6.10. Es sei $u \in H_0^1(\Omega) \cap H^{m+1}(\Omega)$ die L"osung der Laplace-Gleichung zu $f \in H^{m-1}(\Omega)$. Dann gilt f"ur $u_h \in V_h := V_h^{(r),dg}$ als L"osung von

$$a_h(u_h, \phi_h) = (f, \phi_h) \quad \forall \phi_h \in V_h,$$

wobei $a_h(\cdot,\cdot)$ durch (6.22) definiert ist die a priori Fehlerabsch"atzung

$$\|u - u_h\| \leqslant c h^{\min\{r,m\}+1} \|u\|_{H^{m+1}(\Omega)}.$$

Beweis: (i) Wir definieren zun"achst f"ur $e_h := u - u_h$ das duale Problem

$$-\Delta z = \frac{e_h}{\|e_h\|}, \quad (\nabla\phi, \nabla z) = (e_h, \phi)\|e_h\|^{-1} \quad \forall \phi \in H_0^1(\Omega).$$

Da $e_h \in L^2(\Omega)$ gilt bei entsprechender Regularit"at des Gebietes $z \in H^2(\Omega) \cap H_0^1(\Omega)$. Die SIPG-Formulierung ist dual konsistent, denn es gilt f"ur $z \in H_0^1(\Omega)$ mit $[z] = 0$

$$\begin{aligned} a_h(\phi_h, z) &= (\nabla_h \phi_h, \nabla z) + \int_{\Gamma_h} \delta[\phi_h] \underbrace{[z]}_{=0} \,do + \int_{\Gamma_h} \{\{\mathbf{n} \cdot \nabla \phi_h\}\} \underbrace{[z]}_{=0} \,do + \int_{\Gamma_h} [\phi_h] \underbrace{\{\{\mathbf{n} \cdot \nabla z\}\}}_{=\mathbf{n}\cdot\nabla z} \,do \\ &= \sum_{K \in \Omega_h} \Big(-(\phi_h, \Delta z)_K + \int_{\partial K} \phi_h \mathbf{n} \cdot \nabla z \,do - \int_{\partial K} \phi_h \mathbf{n} \cdot \nabla z \,do \Big) \\ &= (e_h, \phi_h)\|e_h\|^{-1}. \end{aligned}$$

F"ur $\phi \in H_0^1(\Omega)$ gilt ebenso $a_h(\phi, z) = (\nabla\phi, \nabla z)$. Wir definieren entsprechend die diskrete duale L"osung $z_h \in V_h$ gem"a"s

$$a_h(\phi_h, z_h) = (e_h, \phi_h)\|e_h\|^{-1} \quad \forall \phi_h \in V_h.$$

(ii) Dann gilt bei Ausnutzung der dualen Konsistenz sowie Orthogonalit"at und Stetigkeit von $a_h(\cdot,\cdot)$

$$\|e_h\| = (\nabla e_h, \nabla z) = a_h(e_h, z) = a_h(e_h, z - i_h z) \leqslant c |||u - u_h|||_\delta |||z - i_h z|||_\delta,$$

und das gew"unschte Ergebnis folgt mit einer Interpolationsabsch"atzung f"ur $z - i_h z$ bei Beachtung von $z \in H^2(\Omega)$. □

F"ur das SIPG-Verfahren erhalten wir also "ahnliche Resultate zu dem "ublichen stetigen Finite Elemente Verfahren. Die Verwendung von dG-Verfahren f"ur elliptische Gleichungen ist nur in Spezialf"allen sinnvoll. Denn im Ergebnis erhalten wir die gleiche numerische Approximationseigenschaft wie bei Verwendung stetiger Ans"atze. Der Aufwand ist jedoch wesentlich h"oher, da entlang von Gitterkanten und in Gitterknoten die Anzahl der Freiheitsgrade vervielfacht ist. Dies spiegelt sich gegebenenfalls in einer g"unstigeren Fehlerkonstante wieder. Bei Ans"atzen erster Ordnung, also im Fall $dG(1)$ hat das dG-Verfahren auf einem regul"aren Viereckgitter die vierfache Anzahl von Freiheitsgraden. Dieser Unterschied wird bei h"oheren Ansatzgraden mit vielen inneren Freiheitsgraden geringer. dG-Verfahren zeichnen sich weiter durch eine hohe Flexibilit"at aus. So k"onnen in verschiedenen Gitterzellen verschiedene Polynomgrade gew"ahlt werden. Diese h/p-Methode der Finiten Elemente kann bei optimal entworfenen Gittern und lokalen Polynomgraden zu sehr guter Konvergenz f"uhren.

6.2.2 Unstetige Galerkin-Verfahren f"ur Transport-Reaktions-Gleichungen

Wir n"ahern uns nun den Euler-Gleichungen und betrachten nun ein reines Konvektionsproblem der Art

$$a u + \nabla \cdot (\beta u) = f, \tag{6.25}$$

wobei $\beta : \Omega \to \mathbb{R}^2$ ein hinreichend glattes Transportfeld ist und $a \geqslant 0$ ein gegebener Parameter. Wir k"onnen dieses Problem alternativ schreiben als

$$\beta \cdot \nabla u + \gamma u = f, \quad \gamma := a + \operatorname{div} \beta.$$

Ausgehend von einem Punkt $x \in \Omega$ betrachten wir nun die L"osung $x(s) \subset \Omega$ der Anfangswertaufgabe

$$x'(s) = \beta(x), \quad x(0) = x.$$

Angenommen, β sei ein Lipschitz-stetiges Feld, so existiert eine eindeutig bestimmte L"osung. Wir nennen diese L"osungen *Charakteristiken* der Transportgleichung (6.25). Es gilt:

$$\frac{d}{ds} u(x(s)) = \sum_{i=1}^{d} \partial_i u \partial_s x_i(s) = \beta \cdot \nabla u,$$

und also

$$\frac{d}{ds} u(x(s)) + \gamma u(x(s)) = 0.$$

Entlang einer Charakteristik ist die L"osung der partiellen Differentialgleichung (6.25) durch die L"osung einer Anfangswertaufgabe bestimmt. Ist $u(x(0))$ auf einem Punkt der Charakteristik bestimmt, so ist unmittelbar die gesamte L"osung auf der Charakteristik vorgegeben

$$u(x(s)) = u(x(0)) + \int_0^s \gamma u(x(t))\, dt.$$

Hieraus k"onnen wir R"uckschl"usse "uber die vorzuschreibenden Randwerte ziehen: auf einem Punkt $x \in \partial\Omega$ mit $\beta \cdot \mathbf{n} < 0$, d.h. auf Randpunkten f"ur die $\beta(x)$ in das Gebiet zeigt, muss ein Dirichlet-Wert f"ur die L"osung vorgegeben werden. Dieser Wert gibt dann die L"osung entlang der gesamten Charakteristik vor. Wir definieren

$$\Gamma_- := \{x \in \partial\Omega,\ \beta \cdot \mathbf{n} < 0\},$$

als den Randanteil, auf dem Dirichlet-Werte vorzugeben sind. Entsprechend sei Γ_+ als der Rand gegebenen, wo die Charakteristiken das Gebiet verlassen. Hier d"urfen keine Randwerte vorgegeben werden. Es ist m"oglich, dass sich Charakteristiken im Gebiet schneiden. In diesem Fall kann die L"osung u Unstetigkeiten aufweisen. Ebenso werden unstetige Randdaten auf Γ_- unstetig im Gebiet fortgesetzt. Das reine Transportproblem hat keine Gl"attungseigenschaft.

Es gilt:

Satz 6.11. Es sei $f \in L^2(\Omega)$ und $a \in L^\infty(\Omega)$ sowie $\beta \in [W^{1,\infty}(\Omega)]^2$ mit

$$a(x) + \frac{1}{2}\nabla \cdot \beta(x) = c_0^2(x) \geqslant \gamma_0 > 0.$$

Dann gilt f"ur eine L"osung von (6.25)

$$\gamma_0 \|u\| \leqslant \|f\|.$$

Beweis: Es gilt bei Multiplikation von (6.25) mit u und Integration "uber Ω

$$(au, u) + (\nabla \cdot (\beta u), u) = (f, u).$$

F"ur den Transportterm gilt

$$(\nabla \cdot (\beta u), u) = -(\beta \cdot \nabla u, u) + \int_{\partial\Omega} \mathbf{n} \cdot \beta |u|^2 \, do.$$

Auf Γ_- sind Dirichlet-Randwerte vorgeschrieben und auf Γ_+ gilt $\mathbf{n} \cdot \beta \geqslant 0$, also folgt mit $\nabla \cdot (\beta u) = \beta \cdot \nabla u + (\nabla \cdot \beta) u$:

$$(\beta \cdot \nabla u, u) + \frac{1}{2}((\nabla \cdot \beta) u, u) \geqslant 0.$$

So gilt

$$(au + \nabla \cdot (\beta u), u) = \left((a + \frac{1}{2}\nabla \cdot \beta) u, u\right) + \underbrace{(\beta \cdot \nabla u, u) + \frac{1}{2}((\nabla \cdot \beta) u, u)}_{\geqslant 0} \geqslant \gamma_0 \|u\|^2.$$

Somit folgt auch $\gamma_0 \|u\| \leqslant \|f\|$. □

Diese Aussage beschr"ankt die L"osung lediglich in der L^2-Norm. Da wir bereits argumentiert haben, dass eine L"osung der Transportgleichung Unstetigkeiten aufweisen kann, ist diese Aussage zun"achst optimal. Im Fall exakt divergenzfreier Transportfelder $\nabla \cdot \beta = 0$ vereinfacht sich das System zu

$$au + \beta \cdot \nabla u = f.$$

In diesem Fall gilt f"ur den Transportterm

$$(\beta \cdot \nabla u, u) = -(\beta \cdot \nabla u, u) + \int_{\Gamma_+} \underbrace{\mathbf{n} \cdot \beta}_{\geqslant 0} |u|^2 \, do \quad \Rightarrow \quad (\beta \cdot \nabla u, u) \geqslant 0,$$

so dass $\alpha > 0$ beliebig gew"ahlt werden kann.

Zur Herleitung einer dG-Formulierung im Raum $V_h := V_h^{(r),dg}$ starten wir wieder mit der klassischen Formulierung. Es gilt f"ur $\phi \in V_h$

$$(f, \phi_h) = \sum_{K \in \Omega_h} (au, \phi_h)_K + (\nabla \cdot (\beta u), \phi_h)_K = \sum_{K \in \Omega_h} (au, \phi_h)_K - (\beta u, \nabla \phi_h)_K + \int_{\partial K} (\mathbf{n} \cdot \beta u, \phi_h)_{\partial K}.$$

Zur Definition einer diskreten Formulierung f"uhren wir erneut eine numerische Flussfunktion $h(\cdot)$ ein:

$$a_h(u_h, \phi_h) := (a u_h, \phi_h) - (\beta u_h, \nabla_h \phi_h) - \sum_{K \in \Omega} \int_{\partial K} h(u_h^+, u_h^-, \mathbf{n}, \beta) \phi_h \, do.$$

Definition 6.3 und Satz 6.4 gelten sinngem"a"s.

Wir kommen nun wieder zu einer Beschreibung der Flussfunktion. Zun"achst w"ahlen wir den naheliegenden Ansatz

$$h(u_h^+, u_h^-, \mathbf{n}, \beta) = \{\{\mathbf{b} \cdot \beta u_h\}\}.$$

F"ur glatte L"osungen u ist dieser Fluss und somit auch die variationelle Formulierung konsistent. Wir gehen nun auch auf die Randwerte an Γ_- sowie Γ_+ ein. Hier definieren wir

$$h(u_h^+, u_h^-, \mathbf{n}, \beta)\Big|_{\Gamma_+} := \mathbf{n} \cdot \beta u_h^-, \quad h(u_h^+, u_h^-, \mathbf{n}, \beta)\Big|_{\Gamma_-} := \mathbf{n} \cdot \beta g,$$

wobei g die auf Γ_- vorgegebenen Randwerte beschreibt.

Bei Transportgleichungen spielen Erhaltungseigenschaften eine wichtige Rolle: bei Abwesenheit externe Einfl"usse wird erwartet, dass der Fluss in ein Volumen dem Fluss aus dem Volumen entspricht. Wir betrachten wieder die exakte L"osung. Zu einer Zelle $K \in \Omega_h$ definieren wir

$$\phi_h^K(x) := \begin{cases} 1 & x \in K \\ 0 & x \notin K \end{cases}.$$

Dann gilt f"ur $a = 0$

$$\int_K f \, dx = -\int_{\partial K} \mathbf{n} \cdot \beta u \, do = \int_{\partial K_-} \underbrace{-\mathbf{n} \cdot \beta}_{\geqslant 0} u \, do - \int_{\partial K_+} \underbrace{\mathbf{n} \cdot \beta}_{\geqslant 0} u \, do.$$

Die Differenz zwischen Einfluss und Ausfluss entspricht stets der Produktion f. Wir definieren

Definition 6.12. Eine Flussfunktion hei"st *konservativ*, falls gilt

$$h(u_h^+, u_h^-, \mathbf{n}, \beta) = -h(u_h^-, u_h^+, -\mathbf{n}, \beta).$$

Es gilt:

Satz 6.13. Es sei $h(\cdot)$ eine konsistente Flussfunktion. Dann ist eine Diskretisierung im Fall $a = 0$ erhaltend genau dann, wenn der Fluss konservativ ist.

BEWEIS: Es sei wieder ϕ_h^K wie oben. Dann gilt

$$\int_K f\,dx = -\int_{\partial K} h(u_h^+, u_h^-, \mathbf{n}, \beta)\,do = -\int_{\partial K_-} h(u_h^+, u_h^-, \mathbf{n}, \beta)\,do - \int_{\partial K_+} h(u_h^+, u_h^-, \mathbf{n}, \beta)\,do.$$

Angenommen, der Fluss ist konservativ. Dann gilt:

$$\int f\,dx = -\int_{\partial K_-} h(u_h^+, u_h^-, \mathbf{n}, \beta)\,do + \int_{\partial K_+} h(u_h^-, u_h^+, -\mathbf{n}, \beta)\,do,$$

wobei

$$\int_{\partial K_+} h(u_h^-, u_h^+, -\mathbf{n}, \beta)\,do, = -\int_{\partial K'_-} h(u_h^+, u_h^-, \mathbf{n}, \beta)\,do,$$

mit der Nachbarzelle K' ist. D.h., der Fluss durch eine Zelle entspricht gerade der Produktion f. □

Der Mittelwert-Fluss Wie im Fall der Laplace-Gleichung definieren wir zun"achst den Fluss als Mittelwert

$$h(u_h^+, u_h^-, \mathbf{n}, \beta) = \{\{\mathbf{n} \cdot \beta u_h\}\} = \frac{1}{2}\left(\mathbf{n} \cdot \beta(u_h^+ + u_h^-)\right).$$

Es ist einfach nachzuweisen, dass dieser Fluss sowohl konsistent als auch konservativ ist. Zur weiteren Analyse werden wir zun"achst Koerzivit"at und Stabilit"at in einer geeigneten Norm herleiten. Es gilt:

Satz 6.14. Der dG-Diskretisierung der Konvektionsgleichung mit Mittelwert-Fluss ist koerziv

$$a_h(u_h, u_h) \geqslant \gamma \|u_h\|^2,$$

und jede L"osung h"angt stetig von den Daten ab

$$\|u_h\| \leqslant \gamma^{-1}\|f\|. \tag{6.26}$$

BEWEIS: (i) Wir zeigen zun"achst die Koerzivit"at: Elementweise gilt:

$$\begin{aligned}
-(\beta u_h, \nabla u_h)_K &= -\frac{1}{2}\int_K \beta \cdot \nabla u_h^2\,dx \\
&= -\frac{1}{2}\int_K \nabla \cdot (\beta u_h^2)\,dx + \frac{1}{2}\int_K (\nabla \cdot \beta) u_h^2\,dx \\
&= -\frac{1}{2}\int_{\partial K} \beta \cdot \mathbf{n}(u_h^-)^2\,do + \frac{1}{2}(\nabla \cdot \beta) u_h^2\,dx.
\end{aligned}$$

Weiter gilt auf einer Kante

$$\begin{aligned}
-\int_e \{\{\mathbf{n} \cdot \beta u_h\}\}[u_h]\,do &= -\int_e \frac{1}{2}\mathbf{n}^- \cdot \beta(u_h^+ + u_h^-)(u_h^+ - u_h^-)\,do \\
&= -\int_e \frac{1}{2}\mathbf{n}^- \cdot \beta\left((u_h^+)^2 - (u_h^-)^2\right)do = \frac{1}{2}\int_e \mathbf{n}^- \cdot \beta(u_h^-)^2\,do - \frac{1}{2}\int_e \mathbf{n}^+ \cdot \beta(u_h^+)^2\,do.
\end{aligned}$$

Da jede Kante doppelt vorkommt folgt

$$-\int_{\Gamma_h} \{\{nt \cdot \beta u_h\}\}[u_h]\,do = \int_{\partial K} \frac{1}{2} \int_{\partial K} \mathbf{n} \cdot \beta (u_h^-)^2\,do.$$

Hiermit gilt

$$a_h(u_h, u_h) = \big(\underbrace{(a + \frac{1}{2}(\nabla \cdot \beta))}_{=c_0^2} u_h, u_h \big) \geqslant \gamma \|u_h\|^2.$$

(*ii*) Jetzt sei $u_h \in V_h$ eine variationelle L"osung von $a_h(u_h, \phi_h) = (f, \phi_h)$. Dann folgt unmittelbar

$$\gamma \|u_h\|^2 = a_h(u_h, u_h) = (f, u_h) \leqslant \|f\|\,\|u_h\|.$$

□

Die Stabilit"atsaussage (6.27) beinhaltet keine Aussagen "uber die globale Regularit"at, die st"uckweise definierte L"osung kann von Zelle zu Zelle beliebig schwanken. Dieses Verhalten zeigt sich auch numerisch und gerade bei L"osungen mit geringer Regularit"at ist der Mittelwert als Fluss nicht brauchbar.

Upwind-Fluss Zum Erreichen einer stabilen Diskretisierung schreiben wir f"ur den numerischen Fluss eine ausgezeichnete Richtung vor. Der *Upwind-Fluss* ist gegeben durch

$$h(u_h^+, u_h^-, \mathbf{n}, \beta) := \begin{cases} \beta \cdot \mathbf{n} u_h^+ & \beta \cdot \mathbf{n} < 0, \\ \beta \cdot \mathbf{n} u_h^- & \beta \cdot \mathbf{n} \geqslant 0, \end{cases}$$

und schaut zur"uck *gegen die Str"omung*. Der Upwind und der Mittelwertfluss lassen sich generisch zu einem Fluss vereinigen:

$$h(u_h^+, u_h^-, \mathbf{n}, \beta) := \{\{\mathbf{n} \cdot \beta u_h\}\} - b_0 [u_h].$$

F"ur $b_0 = 0$ liegt der Mittelwertfluss vor, f"ur $b_0 = \frac{1}{2}|\beta \cdot \mathbf{n}|$ gilt:

$$h(u_h^+, u_h^-, \mathbf{n}, \beta) := \frac{1}{2} \mathbf{n} \cdot \beta (u_h^+ + u_h^-) - \frac{1}{2} |\beta \cdot \mathbf{n}| (u_h^+ - u_h^-) = \begin{cases} \beta \cdot \mathbf{n} u_h^+ & \beta \cdot \mathbf{n} < 0, \\ \beta \cdot \mathbf{n} u_h^- & \beta \cdot \mathbf{n} \geqslant 0, \end{cases}$$

Es gilt entsprechend:

Satz 6.15. Der dG-Diskretisierung der Konvektionsgleichung mit generischem Fluss ist koerziv

$$a_h(u_h, u_h) \geqslant |||u_h|||^2,$$

und jede L"osung h"angt stetig von den Daten ab

$$\|u_h\| \leqslant |||u_h|||. \tag{6.27}$$

Dabei ist die generische Norm gegeben durch:

$$|||u_h|||^2 := \gamma \|u_h\|^2 + \int_{\Gamma_h} b_0 [u_h]^2\,do.$$

F"ur $b_0 = \frac{1}{2}|\beta \cdot \mathbf{n}|$ beinhaltet die Norm Kontrolle "uber die Kantenspr"unge. Eine diskrete L"osung kann also nicht mehr beliebig von Zelle zu Zelle springen. Dieser st"arkere Stabilit"atsbegriff zeigt sich auch in numerischen Rechnungen. Bei Wahl von $b_0 = \frac{1}{2}|\beta \cdot \mathbf{n}|$ beinhaltet die Norm keine Kontrolle "uber Kanten, deren Normalvektor $\mathbf{n}$ orthogonal auf der Flussrichtung β stehen. Dies ist auch erw"unscht, da Unstetigkeiten im Gebiet transportiert werden sollen.

F"ur das weitere Vorgehen ben"otigen wir den folgenden Hilfsatz:

Hilfsatz 6.16 (L^2-Projektion)**.** Es sei $V_h := V_h^{(r),dg}$. Die L^2-Projektion $P_h : L^2(K) \to V_h$ von u ist eindeutig definiert

$$(P_h u - u, \phi_h) = 0 \quad \forall \phi_h \in V_h,$$

und f"ur $u \in H^{m+1}(K)$ mit $m \geqslant r$ gelten die Fehlerabsch"atzungen

$$\|\nabla^k (u - P_h u)\|_K \leqslant h^{r+1-k} \|\nabla^{m+1} u\|_K, \quad 0 \leqslant k \leqslant m,$$

sowie

$$\|u - P_h u\|_{\partial K} \leqslant h^{r+\frac{1}{2}} \|\nabla^{m+1} u\|_K,$$

und im Fall $u \in W^{m+1,\infty}(K)$

$$\|\nabla^K (u - P_h u)\|_{L^\infty(K)} \leqslant c h^{r+1-k} \|\nabla^{m+1} u\|_{L^\infty(K)}.$$

Beweis: (*i*) Die eindeutige Existenz der L^2-Projektion kann auf das L"osen eines linearen Gleichungssystems zur"uckgef"uhrt werden.

(*ii*) F"ur die L^2-Projektion gilt die Orthogonalit"atsbeziehung

$$\|u - P_h u\|^2 = (u - P_h u, u - P_h u) = (u - P_h u, u - \phi_h) \leqslant \|u - P_h u\| \, \|u - \phi_h\|,$$

f"ur beliebige $\phi_h \in V_h$. Die Fehlerabsch"atzungen k"onnen somit auf bekannte Interpolationsabsch"atzungen zur"uckgef"uhrt werden. □

Hiermit k"onnen wir den folgenden zentralen Satz "uber die Approximation der dG-Verfahren beweisen:

Satz 6.17 (Fehlerabsch"atzung f"ur dG-Formulierungen der Konvektionsgleichung)**.** Es sei $u \in H^{m+1}(\Omega)$ und $u_h \in V_h^{(r),dg}$ mit $r \leqslant m$ die diskrete dG-L"osung zu

$$a_h(u_h, \phi_h) = (f, \phi_h) \quad \forall \phi_h \in V_h.$$

Weiter sei $\beta \in W^{1,\infty}(\Omega)^2$. Dann gilt im Fall $b_0 = 0$ (Mittelwert-Fluss)

$$\|u - u_h\| \leqslant C h^r \|u\|_{H^{m+1}(\Omega)},$$

und im Fall $b_0 = \frac{1}{2}|\beta \cdot \mathbf{n}|$ (Upwind-Fluss)

$$|||u - u_h||| \leqslant C h^{r+\frac{1}{2}} \|u\|_{H^{m+1}(\Omega)}.$$

Beweis: (*i*) Wir f"ugen gem"a"s Hilfsatz 6.16 die L^2-Projektion ein

$$u - u_h = \underbrace{u - P_h u}_{=:\eta} - \underbrace{u_h - P_h u}_{=:\xi_h}.$$

Der Projektionsfehler kann entsprechend Hilfsatz 6.16 abgesch"atzt werden. Es gilt (bei Beachtung von $r \leqslant m$):

$$|||\eta|||^2 = |||u - P_h u|||^2 \leqslant c h^{2r+2} \|\nabla^{r+1} u\|^2 + c b_0 h^{2r+1} \|\nabla^{r+1} u\|^2.$$

Dabei entf"allt der zweite Anteil im Fall $b_0 = 0$.

(*ii*) F"ur $\xi_h \in V_h^{(r),dg}$ gilt:

$$|||\xi_h|||^2 = a_h(i_h u - u_h, \xi_h) = \underbrace{a_h(u - u_h, \xi_h)}_{=0} + a_h(i_h u - u, \xi_h)$$

F"ur den verbleibenden Term gilt

$$a_h(\eta, \xi_h) = (\eta, c\xi_h - \beta \cdot \nabla \xi_h) - \int_{\Gamma_h} \{\{\mathbf{n} \cdot \beta\eta\}\}[\xi_h] \, do + \int_{\Gamma_h} b_0 [\eta][\xi_h] \, do.$$

Wir untersuchen die einzelnen Terme. Zun"achst gilt

$$(\eta, c\xi_h) + \int_{\Gamma_h} b_0 [\eta][\xi_h] \, do \leqslant 2|||\eta||| \, |||\xi_h|||.$$

F"ur den verbleibenden Term $-(\eta, \beta \cdot \nabla \xi_h)$ beobachten wir zun"achst, dass $\nabla \xi_h \in V_h^{(r-1),dg}$ und also

$$\bar{\beta} \cdot \nabla \xi_h \in V_h^{(r),dg},$$

wobei $\bar{\beta} := P_h^{(0)} \beta$ die L^2-Projektion von β in den Raum der st"uckweise Konstanten ist. Mit der Orthogonalit"at der L^2-Projektion folgt

$$-(\eta, \beta \cdot \nabla \xi_h) = -(\eta, (\beta - \bar{\beta}) \cdot \nabla \xi_h) \leqslant \sum_K \|\eta\|_K \, \|\beta - \bar{\beta}\|_{L^\infty(K)} \, \|\nabla \xi_h\|_K.$$

Mit der L^∞-Absch"atzung aus Hilfsatz 6.16 sowie der inversen Ungleichung folgt

$$-(\eta, \beta \cdot \nabla \xi_h)_K \leqslant c \|\beta\|_{W^{1,\mathrm{infty}}(K)} \|\eta\|_K \|\xi_h\|_K,$$

und somit

$$(\eta, \beta \cdot \nabla \xi_h)_K \leqslant c \|\beta\|_{W^{1,\infty}} |||\eta||| \, |||\xi_h|||.$$

(*iii*) Schlie"slich bleibt die Betrachtung des Mittelwerts. Wir unterscheiden die F"alle $b_0 = \frac{1}{2}|\beta \cdot \mathbf{n}| > 0$ und $b_0 = 0$. Im Fall $b_0 > 0$ gilt

$$\int_{\Gamma_h} \{\{\mathbf{n} \cdot \beta\eta\}\}[\xi_h] \, do \leqslant \left(\int_{\Gamma_h} b_0^{-1} \{\{\mathbf{n} \cdot \beta\eta\}\}^2 \, do \right)^{\frac{1}{2}} \underbrace{\left(\int_{\Gamma_h} b_0 [\xi_h]^2 \, do \right)^{\frac{1}{2}}}_{\leqslant |||\xi_h|||}.$$

Nun ist mit Hilfsatz 6.16:

$$\int_e b_0^{-1}\{\{\mathbf{n}\cdot\beta\eta\}\}^2\,do = 2\int_e \frac{|\mathbf{n}\cdot\beta|^2}{|\mathbf{n}\cdot\beta|}\{\{\eta\}\}^2\,do \leqslant ch^{2r+1}\|\mathbf{n}\cdot\beta\|_{L^\infty(K)}\|\nabla^{r+1}u\|_K^2.$$

Wir betrachten nun den Fall $b_0 = 0$:

$$\int_{\Gamma_h}\{\{\mathbf{n}\cdot\beta\eta\}\}[\xi_h]\,do \leqslant \left(\int_{\Gamma_h}|\beta\cdot\mathbf{n}|^2\{\{\eta\}\}^2\,do\right)^{\frac{1}{2}}\left(\int_{\Gamma_h}[\xi_h]^2\,do\right)^{\frac{1}{2}}.$$

Der Interpolationsfehler kann wie oben abgesch"atzt werden. Der Sprung $[\xi_h]$ hingegen kann nicht in der Norm $|||\cdot|||$ absorbiert werden. Stattdessen folgern wir mit der Spurabsch"atzung aus Hilfsatz 6.6 und der inversen Ungleichung

$$\sum_{e\in\Gamma_h}\|[\xi_h]\|_e^2 \leqslant c\sum_{K\in\Omega_h} h_K^{-1}\|\xi_h\|_K^2 + \|\xi_h\|_K\,\|\nabla\xi_h\|_K \leqslant c\sum_{K\in\Omega_h} h_K^{-1}\|\xi_h\|_K^2.$$

Zusammen gilt

$$\int_{\Gamma_h}\{\{\mathbf{n}\cdot\beta\eta\}\}[\xi_h]\,do \leqslant ch^r\|\mathbf{n}\cdot\beta\|_\infty^2\|\nabla^{r+1}u\|\|\xi_h\|.$$

An dieser Stelle verlieren wir eine halbe h-Potenz. □

Ein Vergleich der beiden Fehlerabsch"atzungen zeigt sofort eine verbesserte (optimale) Konvergenzordnung im Fall des Upwind-Flusses. Insbesondere f"ur $r = 0$ liefert das Upwind-Verfahren eine konvergente L"osung, w"ahrend der Mittelwert-Fluss keine Konvergenz garantiert.

Randwerte Schlie"slich betrachten wir noch die Realisierung von Randwerten $u = g$ auf dem Einstr"omrand Γ_-. Im dG-Sinne werden diese nicht im starken Sinne gefordert, sondern variationell mit Hilfe eines geeigneten Flusses realisiert. Auf Kanten $e \in \Gamma_h \cap \partial\Omega$ des Randes modifizieren wir die Flussfunktion zu

$$h(u_h^+, u_h^-, \beta, \mathbf{n}, g) = \begin{cases} \beta\cdot\mathbf{n}g & \text{auf } \Gamma_-, \\ \beta\cdot\mathbf{n}u_h^- & \text{auf } \Gamma_+ \end{cases},$$

wobei g die vorgeschriebenen Randwerte sind. Schlie"slich k"onnen wir die variationelle Formulierung mit generischem Fluss konkretisieren zu

$$\begin{aligned} a_h(u_h,\phi_h) = (au_h,\phi_h) - (\beta u_h,\nabla_h\phi_h) - \int_{\Gamma_h\setminus\partial\Omega}\{\{\mathbf{n}\cdot\beta u_h\}\}[\phi_h] + b_0[u_h][\phi_h]\}\,do \\ + \int_{\Gamma_h\cap\Gamma_-}\mathbf{n}\cdot\beta g\phi_h\,do + \int_{\Gamma_h\cap\Gamma_+}\mathbf{n}\cdot\beta g u_h^-\,do. \end{aligned}$$

Ein Modellproblem Wir konkretisieren beide dG-Verfahren zur Diskretisierung des eindimensionalen Modellproblems auf $I = [0,1]$. Es sei

$$i = 0,\dots,N: \quad x_i = ih, \quad I_n := (x_{n-1}, x_n), \quad h = \frac{1}{N},$$

und die diskrete L"osung im Raum der st"uckweise Konstanten gegeben als

$$u_h \in V_h^{(0),dg}, \quad u_h\big|_{I_n:=(x_{n-1},x_n]} =: u^n \in \mathbb{R}.$$

Wir konkretisieren nun beide Varianten des numerischen Flusses f"ur das einfache Testproblem

$$cu + \partial_x u = f.$$

Zun"achst betrachten wir den Mittelwertfluss. Es ist (bei Beachtung aller inneren Kanten $x_1,\dots,x_{N-1}$):

$$\begin{aligned}
a_h(u_h,\phi_h) &= \sum_{n=1}^{N} c h u_h^n \phi_h^n - \frac{1}{2}\sum_{n=1}^{N-1}(u_h^n + u_h^{n+1})(\phi_h^{n+1} - \phi_h^n) \\
&= \sum_{n=1}^{N} c h u_h^n \phi_h^n - \frac{1}{2}\sum_{n=2}^{N}(u_h^{n-1} + u_h^n)\phi_h^n + \frac{1}{2}\sum_{n=1}^{N-1}(u_h^n + u_h^{n+1})\phi_h^n \\
&= \sum_{n=1}^{N} c h u_h^n \phi_h^n + \frac{1}{2}\sum_{n=2}^{N-1}(u_h^{n+1} - u_h^{n-1})\phi_h^n - \frac{1}{2}(u_h^{N-1} + u_h^N)\phi_h^N + \frac{1}{2}(u_h^1 + u_h^2)\phi_h^1
\end{aligned}$$

Somit ergibt sich im Innern des Gebietes ein lineares Gleichungssystem

$$c u_h^n + \frac{u_h^{n+1} - u_h^{n-1}}{2h} = f_h^n, \quad n = 2,\dots,N-1.$$

Dies entspricht gerade einer Diskretisierung mit zentralen Finiten Differenzen. Der Zentrale Differenzenquotient ist - wie schon in Kapitel **??** gesehen - nicht stabil zur Diskretisierung von transportdominanten Prozessen.

Im Anschluss betrachten wir nun alternativ den Upwind-Fluss. Hier gilt bei $\beta = 1$ und $\beta \cdot \mathbf{n} = 1$

$$\begin{aligned}
a_h(u_h,\phi_h) &= \sum_{n=1}^{N} c h u_h^n \phi_h^n - \sum_{n=1}^{N-1} u_h^n(\phi_h^{n+1} - \phi_h^n) \\
&= \sum_{n=1}^{N} c h u_h^n \phi_h^n - \sum_{n=2}^{N} u_h^{n-1}\phi_h^n + \sum_{n=1}^{N-1} u_h^n \phi_h^n \\
&= \sum_{n=1}^{N} c h u_h^n \phi_h^n + \sum_{n=2}^{N-1}(u_h^n - u_h^{n-1})\phi_h^n - u_h^{N-1}\phi_h^N + u_h^1\phi_h^1.
\end{aligned}$$

Hier ist die L"osung u_h gegeben als L"osung eines linearen Gleichungssystems

$$c u_h^n + \frac{u_h^n - u_h^{n-1}}{h} = f_h^n,$$

welches einer stabilen Finite Differenzen Approximation mit r"uckw"artigem Differenzenquotienten entspricht.

Der wesentliche Vorteil von dG-Verfahren - auch zur stabilen Diskretisierung von transportdominanten elliptischen Problemen - ist das klare Konstruktionsprinzip, welches sich sofort auf beliebige Ordnungen und auch auf lokal verfeinerte Gitter "ubertragen l"asst.

6.2.3 Unstetige Galerkin Verfahren f"ur die Euler-Gleichungen

Schlie"slich wird die Idee der dG-Verfahren auf die station"aren Euler-Gleichungen

$$\nabla \cdot \begin{pmatrix} \rho \mathbf{v} \\ \rho \mathbf{v} \otimes \mathbf{v} + p\mathrm{I} \\ \rho E \mathbf{v} + p\mathbf{v} \end{pmatrix} = 0,$$

angewendet. Geschlossen wird das System durch ein Gasgesetz, etwa

$$p = (\gamma - 1)\rho e,$$

mit $\gamma = c_p/c_v$ und der inneren Energie e, welche zur totalen Energie in Beziehung steht durch

$$E = e + \frac{1}{2}|\mathbf{v}|^2.$$

Wir f"uhren zur k"urzeren Schreibweise die Bezeichnungen

$$\mathbf{u} := \begin{pmatrix} \rho \\ \rho \mathbf{v} \\ \rho E \end{pmatrix}, \quad F(\mathbf{u}) := \begin{pmatrix} \rho \mathbf{v} \\ \rho \mathbf{v} \otimes \mathbf{v} + p\mathrm{I} \\ \rho E \mathbf{v} + p\mathbf{v} \end{pmatrix},$$

ein. Gesucht wird also - formal sehr "ahnlich zur Transportgleichung - die L"osung $\mathbf{u}$ von

$$\nabla \cdot F(\mathbf{u}) = 0,$$

mit einem $F(\mathbf{u}) \in \mathbb{R}^{(2+d)\times 3}$. Zum Herleiten einer diskreten Formulierung definieren wir zun"achst den diskreten L"osungsraum

$$V_h := [V_h^{(r),dg}]^{2+d},$$

f"ur die Dichte ρ, die d Komponenten des Impulses $\rho\mathbf{v}$ sowie die Energiedichte ρE. Wir w"ahlen hier *equal-order* Finite Elemente f"ur alle L"osungskomponenten. Andere Ans"atze sind denkbar, da wir jedoch keine inf-sup Bedingung zu erf"ullen haben ist dieser einfache Ansatz naheliegend. Nun gilt

$$\begin{aligned}(\nabla \cdot F(\mathbf{u}), \phi_h) &= \sum_{K \in \Omega_h} -\int_K F(\mathbf{u}) : \nabla \phi_h \, dx + \int_{\partial K} F(\mathbf{u})\mathbf{n} \cdot \phi_h \, do \\ &= -(F(\mathbf{u}), \nabla_h u_h) - + \sum_{K \in \Omega_h} \int_{\partial K \setminus \partial\Omega} F(\mathbf{u})\mathbf{n}\phi_h^- \, do + \int_{\partial\Omega} F(\mathbf{u})\mathbf{n}\phi_h^- \, do.\end{aligned}$$

Zur Diskretisierung mit unstetigen Funktionen $\mathbf{u}_h$ f"uhren wir wieder numerische Flussfunktionen $h(\cdot)$ sowie auf dem Rand $h_\Gamma(\cdot)$ ein. So ist:

$$a_h(\mathbf{u}_h, \phi_h) = -(F(\mathbf{u}_h), \nabla_h \phi_h) + \sum_{K \in \Omega_h} \int_{\partial_K \setminus \partial\Omega} h(\mathbf{u}_h^+, \mathbf{u}_h^-, \mathbf{n}) \phi_h^- \, do + \int_{\Gamma_h \cap \partial\Omega} h_\Gamma(\mathbf{u}_h^-, \mathbf{n}) \phi_h^- \, do.$$

Eine Flussfunktion hei"st wieder *konsistent*, falls f"ur glatte L"osungen gilt

$$h(\mathbf{u}, \mathbf{u}, \mathbf{n}) = -F(\mathbf{u})\mathbf{n},$$

und aus der Konsistenz des Flusses folgt sofort die Konsistenz der diskreten variationellen Formulierung $a_h(\cdot,\cdot)$. Ebenso nennen wir die Flussfunktion *konservativ*, falls gilt

$$h(\mathbf{u}_h^+, \mathbf{u}_h^-, \mathbf{n}) = -h(\mathbf{u}_h^-, \mathbf{u}_h^+, -\mathbf{n}),$$

denn aus dieser Definition folgt f"ur $\phi_S = 1$ auf einer Vereinigung von Elementen $S = K_1 \cup \cdots \cup K_n$ und $\phi_S = 0$ sonst da jede Kante doppelt vorkommt:

$$\begin{aligned} 0 = a_h(\mathbf{u}_h, \phi_S) &= \sum_{K \in S} \int_{\partial K \setminus \partial\Omega} h(\mathbf{u}_h^+, \mathbf{u}_h^-, \mathbf{n}) \phi_S^- \, do \\ &= \sum_{e \in S \setminus \partial S} \int_e \big(\underbrace{h(\mathbf{u}_h^+, \mathbf{u}_h^-, \mathbf{n}) + h(\mathbf{u}_h^-, \mathbf{u}_h^+, -\mathbf{n})}_{=0} \big) \, do + \int_{\partial S} h(\mathbf{u}_h^+, \mathbf{u}_h^-, \mathbf{n}) \, do. \end{aligned}$$

Der Fluss "uber eine Vereinigung von beliebigen Elementen S ist somit Null.

F"ur konservative Fl"usse k"onnen wir die variationelle Formulierung wieder schreiben als

$$a_h(\mathbf{u}_h, \phi_h) = -(F(\mathbf{u}_h), \nabla_h \phi_h) - \int_{\Gamma_h \setminus \partial\Omega} h(\mathbf{u}_h^+, \mathbf{u}_h^-, \mathbf{n}) [\phi_h] \, do + \int_{\Gamma_h \cap \partial\Omega} h_\Gamma(\mathbf{u}_h^-, \mathbf{n}) \phi_h^- \, do.$$

Die Wahl eines numerischen Flusses ist bei den Euler-Gleichungen nicht so einfach zu beantworten. Je nach Str"omungssituation - insbesondere in Abh"angigkeit von der lokalen Machzahl - "andert sich der Typ der Gleichungen. Wir wissen, dass L"osungen der Euler-Gleichungen Unstetigkeiten im Gebiet aufweisen k"onnen. An diesen Schocks darf nat"urlich auch numerisch keine Glattheit erzwungen werden. Ist die Str"omungskonfiguration jedoch "uberall subsonisch ($Ma < 1$), so kann auch eine glatte L"osung erwartet werden.

Eine oft verwendete Flussfunktion ist der *Lax-Friedrichs Fluss*

$$h(\mathbf{u}_h^+, \mathbf{u}_h^-, \mathbf{n}) = \frac{1}{2} \left(F(\mathbf{u}_h^+)\mathbf{n} + F(\mathbf{u}_h^-)\mathbf{n} + \alpha[\mathbf{u}] \right).$$

Dabei ist der Parameter α nun nicht mehr ad hoc zu w"ahlen, sondern h"angt von der lokalen Str"omungssituation ab. Es gilt

$$\alpha = \max_{\mathbf{v} \in \{\mathbf{u}_h^+, \mathbf{u}_h^-\}} \{|\lambda(\mathbf{B}(\mathbf{v}, \mathbf{n}))|\},$$

also der maximale Eigenwert der Matrix

$$\mathbf{B}(\mathbf{v},\mathbf{n}) := \nabla_u F(\mathbf{v})\mathbf{n}.$$

Da $F(\cdot)$ matrixwertig ist, ist $\nabla_u F \in \mathbb{R}^{(2+d)\times(2+d)}$ ein Tensor dritter Stufe, $\mathbf{B} := \nabla_u F(\mathbf{v})\mathbf{n} \in \mathbb{R}^{(2+d)\times(2+d)}$ ist wieder ein Tensor zweiter Stufe, kann also als Matrix dargestellt werden. F"ur die Eigenwerte der Matrix $\mathbf{B}(\mathbf{v},\mathbf{n})$ gilt

$$\lambda_{1/2} = \mathbf{v}\cdot\mathbf{n} \pm c, \quad \lambda_{3/4} = \mathbf{v}\cdot\mathbf{n}, \tag{6.28}$$

mit der (lokalen) Schallgeschwindigkeit c. Im Fall $Ma = 1$ kann ein Eigenwert verschwinden, so dass die Matrix $\mathbf{B}$ ihre Regularit"at verliert. F"ur $Ma \approx 1$ ist die Konditionierung der Matrix sehr schlecht. Angewendet auf die lineare Transportgleichung $\nabla\cdot(\beta u) = 0$ entspricht diese Flussfunktion gerade dem Upwind-Fluss.

Randwerte Schlie"slich gehen wir noch kurz auf die Randwerte der Euler-Gleichung ein. Die Annahme der Haftrandbedingung beruht bei den Navier-Stokes Gleichungen auf der Viskosit"at des Fluids, welche daf"ur sorgt, dass sich die Str"omung durch innere Reibung zum Rand hin verlangsamt. Die Euler-Gleichungen beschreiben reibungsfreie Fluide. Hier kann keine Haftrandbedingung erzwungen werden. Stattdessen wird eine - physikalisch sinnvolle - Wandbedingung in Normalrichtung

$$\mathbf{v}\cdot\mathbf{n} = 0 \quad \text{auf } \Gamma_{\text{Wand}},$$

gefordert. L"osungen der Euler-Gleichungen erzeugen keine Grenzschicht und es treten auch keine Turbulenzen auf. Dies unterscheidet die Euler-Gleichung wesentlich von den Navier-Stokes Gleichungen.

Die Vorgabe von Randbedingungen an Einstr"om- oder Ausstr"omr"andern ist aufw"andiger und h"angt wieder von der konkreten Str"omungssituation ab ($Ma < 1$ oder $Ma > 1$). Im Fall einer echt supersonischen $Ma > 1$ Str"omung im Fernfeld ist die Lage klar. Hier hat die Gleichung den Typ eines Transportproblems. Auf dem Einstr"omrand

$$\Gamma_- := \{x \in \partial\Omega,\ \mathbf{v}(x)\cdot\mathbf{n}(x) < 0\},$$

werden Dirichletwerte f"ur alle L"osungskomponenten vorgeschrieben

$$\mathbf{u} = \mathbf{u}^D \text{ auf } \Gamma_-,$$

auf dem Ausstr"omrand m"ussen keine Randwerte vorgegeben werden. In diesem Fall gilt f"ur die Eigenwerte (6.28) $\lambda_i < 0$ auf allen Einstr"omr"andern. Im subsonischen Bereich gilt auf dem Einstr"omrand

$$\lambda_1, \lambda_3, \lambda_4 < 0, \quad \lambda_2 > 0.$$

Hier werden f"ur den Impuls $\rho\mathbf{v}$ sowie die Energie ρE Dirichletwerte auf Γ_-. Der Druck wird "ublicherweise auf dem Ausstr"omrand Γ_+ spezifiziert.