Arnold Reusken

Numerical Methods for the Navier-Stokes equations

January 6, 2012

Chair for Numerical Mathematics
RWTH Aachen

Contents

1 The Navier-Stokes equations 1
1.1 Derivation of the equations 1
1.2 Related models 4
1.3 Weak formulations 7
1.3.1 Function spaces 7
1.3.2 Oseen problem in weak formulation 10
1.3.3 Time dependent (Navier-)Stokes equations in weak formulation 12

2 Finite element semi-discretization of the Navier-Stokes equations 19
2.1 Hood-Taylor finite element spaces 19
2.1.1 Simplicial finite element spaces 19
2.1.2 Hood-Taylor finite element discretization of the Oseen problem 21
2.1.3 Matrix-vector representation of the discrete problem 23
2.1.4 Hood-Taylor semi-discretization of the non-stationary (Navier-)Stokes problem 24
2.2 Numerical experiments 29
2.2.1 Flow in a rectangular tube 29
2.2.2 Flow in a curved channel 30
2.3 Discussion and additional references 31

3 Time integration of semi-discrete Navier-Stokes equations 37
3.1 Introduction 37
3.2 The θ-scheme for the Navier-Stokes problem 43
3.3 Fractional-step θ-scheme for the Navier-Stokes problem 46
3.4 Numerical experiments 46
3.5 Discussion and additional references 48

4 Appendix: Variational formulations in Hilbert spaces 49
4.1 Variational problems and Galerkin discretization 49
4.2 Application to elliptic problems 51
4.3 Application to saddle point problems 52

References 57

1

The Navier-Stokes equations

1.1 Derivation of the equations

We always assume that the physical domain $\Omega \subset \mathbb{R}^3$ is an open bounded domain. This domain will also be the computational domain. We consider the flow problems for a fixed time interval denoted by $[0, T]$. We derive the Navier-Stokes equations for modeling a laminar fluid flow. We assume the fluid to be incompressible, viscous, Newtonian and pure (i. e., no mixture of different components). Moreover we assume isothermal conditions and therefore neglect variations of density and dynamic viscosity due to temperature changes. Hence, dynamic viscosity and, due to incompressibility, also the density are constant (and positive).

The Eulerian coordinates of a point in Ω are denoted by $x = (x_1, x_2, x_3)$. We take a fixed $t_0 \in (0, T)$ and consider a time interval $(t_0 - \delta, t_0 + \delta)$, with $\delta > 0$ sufficiently small such that for $t \in (t_0 - \delta, t_0 + \delta)$ the quantities introduced below are well-defined. Let $\mathbf{X}$ denote a particle (also called "material point") in Ω at $t = t_0$, with Eulerian coordinates $\xi \in \mathbb{R}^3$. Let $X_\xi(t)$ denote the Eulerian coordinates of the particle $\mathbf{X}$ at time t. The mapping

$$t \to X_\xi(t), \quad t \in (t_0 - \delta, t_0 + \delta),$$

describes the trajectory of the particle $\mathbf{X}$. The particles are transported by a velocity field, which is denoted by $\mathbf{u} = \mathbf{u}(x, t) = \big(u_1(x, t), u_2(x, t), u_3(x, t)\big) \in \mathbb{R}^3$. Hence

$$\frac{d}{dt} X_\xi(t) = \mathbf{u}(X_\xi(t), t). \tag{1.1}$$

For the given $\mathbf{X}$, the solution of the system of ordinary differential equations

$$\frac{d}{dt} X_\xi(t) = \mathbf{u}\big(X_\xi(t), t\big), \quad t \in (t_0 - \delta, t_0 + \delta), \quad X_\xi(t_0) = \xi,$$

yields the trajectory of the particle $\mathbf{X}$.

Physical processes can be modeled in different coordinate systems. For flow problems, the two most important ones are (x, t) ("Eulerian") and (ξ, t) ("Lagrangian"):

- Euler coordinates (x, t): one takes an arbitrary fixed point x in space and considers the velocity $\mathbf{u}(x, t)$ at x. If time evolves *different* particles pass through x.
- Lagrange (or "material") coordinates (ξ, t): one takes an arbitrary fixed particle (material point) and considers its motion. If time evolves one thus follows the trajectory of a *fixed* particle.

Related to the Lagrangian coordinates we define the so-called *material derivative* of a (sufficiently smooth) function $f(x, t)$ on the trajectory of $\mathbf{X}$:

$$\dot{f}(X_\xi(t), t) := \frac{d}{dt} f(X_\xi(t), t).$$

If f is defined in a neighborhood of the trajectory we obtain from the chain rule and (1.1):

$$\dot{f} = \frac{\partial f}{\partial t} + \mathbf{u} \cdot \nabla f. \tag{1.2}$$

The derivation of partial differential equations that model the flow problem is based on conservation laws applied on a (small) subdomain, called a material volume, $W_0 \subset \Omega$. We derive these partial differential equations in Eulerian coordinates. Given W_0, define

$$W(t) := \{\, X_\xi(t) : \xi \in W_0 \,\}.$$

$W(t)$ describes the position of the particles at time t, which were located in W_0 at time $t = t_0$. We need the following fundamental identity, which holds for a scalar sufficiently smooth function $f = f(x,t)$:

Reynolds' transport theorem:

$$\begin{aligned} \frac{d}{dt}\int_{W(t)} f(x,t)\,dx &= \int_{W(t)} \dot{f}(x,t) + f \operatorname{div} \mathbf{u}(x,t)\,dx \\ &= \int_{W(t)} \frac{\partial f}{\partial t}(x,t) + \operatorname{div}(f\mathbf{u})(x,t)\,dx, \\ \text{with}\quad \dot{f} &:= \frac{\partial f}{\partial t} + \mathbf{u}\cdot\nabla f \quad \text{the material derivative.} \end{aligned} \tag{1.3}$$

First we consider the *conservation of mass* principle. Let $\rho(x,t)$ be the *density* of the fluid. If we take $f = \rho$ in (1.3) this yields

$$0 = \frac{d}{dt}\int_{W(t)} \rho\,dx = \int_{W(t)} \frac{\partial \rho}{\partial t} + \operatorname{div}(\rho\mathbf{u})\,dx,$$

which holds in particular for $t = t_0$ and for an arbitrary material volume $W(t_0) = W_0$ in Ω. Since also $t_0 \in (0,T)$ is arbitrary, we obtain the partial differential equation

$$\frac{\partial \rho}{\partial t} + \operatorname{div}(\rho\mathbf{u}) = 0 \quad \text{in} \quad \Omega \times (0,T).$$

Due to the assumption $\rho = const$ this simplifies to

$$\operatorname{div} \mathbf{u} = 0 \quad \text{in} \quad \Omega \times (0,T), \tag{1.4}$$

which is often called *mass conservation equation* or *continuity equation*.

We now consider *conservation of momentum*. The momentum of mass contained in $W(t)$ is given by

$$M(t) = \int_{W(t)} \rho\mathbf{u}\,dx.$$

Due to Newton's law the *change* of momentum $M(t)$ is equal to the force $F(t)$ acting on $W(t)$. This force is decomposed in a *volume* force $F_1(t)$ and a *boundary* force $F_2(t)$. We restrict ourselves to the case where the only volume force acting on the volume $W(t)$ is gravity:

$$F_1(t) = \int_{W(t)} \rho\mathbf{g}\,dx,$$

where $\mathbf{g} \in \mathbb{R}^3$ is the vector of gravitational acceleration. The boundary force $F_2(t)$ is used to describe internal forces, i.e., forces that a fluid exerts on itself. These include pressure and the viscous drag that a fluid element $W(t)$ gets from the adjacent fluid. These internal forces are

contact forces: they act on the boundary $\partial W(t)$ of the fluid element $W(t)$. Let $\vec{t}$ denote this internal force vector, also called traction vector. Then we have

$$F_2(t) = \int_{\partial W(t)} \vec{t}\, ds.$$

Cauchy derived fundamental principles of continuum mechanics and in particular he derived the following law (often called *Cauchy's theorem*):

$$\vec{t} \text{ is a linear function of } \mathbf{n},$$

where $\mathbf{n} = \mathbf{n}(x,t) \in \mathbb{R}^3$ is the outer unit normal on $\partial W(t)$. For more explanation on this we refer to introductions to continuum mechanics, for example [22]. Thus it follows that there is a matrix $\boldsymbol{\sigma} = \boldsymbol{\sigma}(x,t) \in \mathbb{R}^{3\times 3}$, called the *stress tensor*, such that the boundary force can be represented as

$$F_2(t) = \int_{\partial W(t)} \boldsymbol{\sigma}\mathbf{n}\, ds. \tag{1.5}$$

Using these force representations in Newton's law and applying Stokes' theorem for $F_2(t)$ we get

$$\begin{aligned} \frac{d}{dt}M(t) &= F_1(t) + F_2(t) \\ &= \int_{W(t)} \rho\mathbf{g} + \operatorname{div}\boldsymbol{\sigma}\, dx. \end{aligned} \tag{1.6}$$

For a matrix $\mathbf{A}(x) \in \mathbb{R}^3$, $x \in \mathbb{R}^3$, its divergence is defined by

$$\operatorname{div}\mathbf{A}(x) = \begin{pmatrix} \operatorname{div}(a_{11}\ a_{12}\ a_{13}) \\ \operatorname{div}(a_{21}\ a_{22}\ a_{23}) \\ \operatorname{div}(a_{31}\ a_{32}\ a_{33}) \end{pmatrix} \in \mathbb{R}^3.$$

Using the transport theorem (1.3) in the left-hand side of (1.6) with $f = \rho\, u_i$, $i = 1,2,3$, we obtain

$$\int_{W(t)} \frac{\partial \rho\, u_i}{\partial t} + \operatorname{div}(\rho\, u_i\, \mathbf{u})\, dx = \int_{W(t)} \rho\, g_i + \operatorname{div}\boldsymbol{\sigma}_i\, dx, \qquad i = 1,2,3,$$

with $\boldsymbol{\sigma}_i$ the i-th row of $\boldsymbol{\sigma}$ and g_i the i-th component of $\mathbf{g}$. In vector notation, with $\mathbf{u}\otimes\mathbf{u} = (u_i u_j)_{1\le i,j\le 3}$,

$$\int_{W(t)} \frac{\partial \rho\,\mathbf{u}}{\partial t} + \operatorname{div}(\rho\,\mathbf{u}\otimes\mathbf{u})\, dx = \int_{W(t)} \rho\,\mathbf{g} + \operatorname{div}\boldsymbol{\sigma}\, dx, \tag{1.7}$$

which holds in particular for $t = t_0$ and for an arbitrary material volume $W(t_0) = W_0$ in Ω. Since $t_0 \in (0,T)$ is arbitrary, we obtain the partial differential equations

$$\frac{\partial \rho\,\mathbf{u}}{\partial t} + \operatorname{div}(\rho\,\mathbf{u}\otimes\mathbf{u}) = \rho\,\mathbf{g} + \operatorname{div}\boldsymbol{\sigma} \quad \text{in} \quad \Omega\times(0,T).$$

Note that $\operatorname{div}(\rho\,\mathbf{u}\otimes\mathbf{u}) = \rho(\mathbf{u}\cdot\nabla)\mathbf{u} + \rho\,\mathbf{u}\operatorname{div}\mathbf{u}$ and due to the continuity equation (1.4), the last summand vanishes, yielding the so-called *momentum equation*

$$\rho\,\frac{\partial \mathbf{u}}{\partial t} + \rho\,(\mathbf{u}\cdot\nabla)\mathbf{u} = \rho\,\mathbf{g} + \operatorname{div}\boldsymbol{\sigma}. \tag{1.8}$$

For viscous *Newtonian fluids* one assumes that the stress tensor $\boldsymbol{\sigma}$ is of the form

$$\boldsymbol{\sigma} = -p\mathbf{I} + L(\mathbf{D}), \tag{1.9}$$

where p is the pressure,

$$\mathbf{D}(\mathbf{u}) = \nabla\mathbf{u} + (\nabla\mathbf{u})^T$$

is the deformation tensor, $\nabla\mathbf{u} := \big(\nabla u_1\ \nabla u_2\ \nabla u_3\big)$, and L is assumed to be a *linear* mapping. Based on this structural model for the stress tensor and using the additional assumptions that the medium is isotropic (i.e. its properties are the same in all space directions) and the action of the stress tensor is independent of the specific frame of reference ("invariance under a change in observer") it can be shown ([22, 17]) that the stress tensor must have the form

$$\boldsymbol{\sigma} = -p\mathbf{I} + \lambda \operatorname{div}\mathbf{u}\,\mathbf{I} + \mu\mathbf{D}(\mathbf{u}). \tag{1.10}$$

Further physical considerations lead to relations for the parameters μ, λ, e.g., $\mu > 0$ (for a viscous fluid), $\lambda \geq -\frac{2}{3}\mu$ or even $\lambda = -\frac{2}{3}\mu$. For the case of an incompressible fluid, i.e., $\operatorname{div}\mathbf{u} = 0$, the relation for the stress tensor simplifies to

$$\boldsymbol{\sigma} = -p\mathbf{I} + \mu\mathbf{D}(u), \tag{1.11}$$

with $\mu > 0$ the *dynamic viscosity*. Hence, we obtain the fundamental Navier-Stokes equations for incompressible flow:

$$\begin{aligned} \rho\big(\frac{\partial\mathbf{u}}{\partial t} + (\mathbf{u}\cdot\nabla)\mathbf{u}\big) &= -\nabla p + \operatorname{div}(\mu\mathbf{D}(\mathbf{u})) + \rho\mathbf{g} \text{ in } \ \Omega \\ \operatorname{div}\mathbf{u} &= 0 \text{ in } \ \Omega. \end{aligned} \tag{1.12}$$

These equations are considered for $t \in [0, T]$. Initial and boundary conditions corresponding to these Navier-Stokes equations are discussed below.

Remark 1.1.1 Using the assumption that μ is a strictly positive *constant* and the relation $\operatorname{div}\mathbf{u} = 0$ we get

$$\operatorname{div}(\mu\mathbf{D}(\mathbf{u})) = \mu\Delta\mathbf{u} = \mu\begin{pmatrix}\Delta u_1\\ \Delta u_2\\ \Delta u_3\end{pmatrix}.$$

Remark 1.1.2 For the Navier-Stokes model one needs suitable initial and boundary conditions only for the velocity $\mathbf{u}$. The initial condition is $\mathbf{u}(x,0) = \mathbf{u}_0(x)$ with a given function $\mathbf{u}_0$, which usually comes from the underlying physical problem. For the boundary conditions we distinguish between *essential* and *natural* boundary conditions. Let $\partial\Omega$ be subdivided into two parts $\partial\Omega = \partial\Omega_D \cup \partial\Omega_N$ with $\partial\Omega_D \cap \partial\Omega_N = \emptyset$. We use essential boundary conditions on $\partial\Omega_D$ that are of Dirichlet type. In applications these describe inflow conditions or conditions at walls (e.g., no-slip). Such Dirichlet conditions are of the form $\mathbf{u}(x,t) = \mathbf{u}_D(x,t)$ for $x \in \partial\Omega_D$, with a given function $\mathbf{u}_D$. If, for example, $\partial\Omega_D$ corresponds to a fixed wall, then a no-slip boundary condition is given by $\mathbf{u}(x,t) = 0$ for $x \in \partial\Omega_D$. On $\partial\Omega_N$ we prescribe natural boundary conditions, which are often used to describe outflow conditions. These natural boundary conditions are of the form

$$\boldsymbol{\sigma}\mathbf{n}_\Omega = -p_{ext}\mathbf{n}_\Omega, \quad \text{on } \ \partial\Omega_N, \tag{1.13}$$

with $\mathbf{n}_\Omega$ the outward pointing normal on $\partial\Omega_N$ and p_{ext} a given function (external pressure). For the case $p_{ext} = 0$ we thus obtain a homogeneous natural boundary condition.

1.2 Related models

We recall the non-stationary Navier-Stokes equations for modeling a *one*-phase incompressible flow problem:

$$\begin{aligned} \rho\big(\frac{\partial \mathbf{u}}{\partial t} + (\mathbf{u}\cdot\nabla)\mathbf{u}\big) + \nabla p - \mu\Delta\mathbf{u} &= \rho\mathbf{g} \text{ in } \ \Omega \\ \operatorname{div}\mathbf{u} &= 0 \text{ in } \ \Omega, \end{aligned} \tag{1.14}$$

with given constants $\rho > 0,\ \mu > 0$. For simplicity we only consider homogeneous Dirichlet boundary conditions for the velocity (no-slip condition). Thus the boundary and initial conditions are given by

$$\mathbf{u} = 0 \quad \text{on } \ \partial\Omega, \quad \mathbf{u}(x,0) = \mathbf{u}_0(x) \quad \text{for } x \in \Omega, \tag{1.15}$$

with a given initial condition $\mathbf{u}_0(x)$. In the discussion and analysis of numerical methods for this problem we will also use the following two simpler systems of partial differential equations. Firstly, the *non-stationary Stokes equations*

$$\begin{aligned} \rho\frac{\partial \mathbf{u}}{\partial t} + \nabla p - \mu\Delta\mathbf{u} &= \rho\mathbf{g} \text{ in } \ \Omega \\ \operatorname{div}\mathbf{u} &= 0 \text{ in } \ \Omega, \end{aligned} \tag{1.16}$$

with the same initial and boundary conditions as in (1.15). Note that opposite to the Navier-Stokes equations, the Stokes system is *linear* in the unknowns $\mathbf{u}$, p. Secondly, we consider the following type of *stationary* problem:

$$\begin{aligned} \xi\mathbf{u} + (\mathbf{w}\cdot\nabla)\mathbf{u} + \nabla p - \mu\Delta\mathbf{u} &= \rho\mathbf{g} \text{ in } \ \Omega \\ \operatorname{div}\mathbf{u} &= 0 \text{ in } \ \Omega, \end{aligned} \tag{1.17}$$

with a given constant $\xi \geq 0$ and a given vector function $\mathbf{w}(x) \in \mathbb{R}^3$. For the boundary condition we take the homogeneous Dirichlet condition $\mathbf{u} = 0$. This linear system of partial differential equations is called an *Oseen problem*. This type of problem occurs if in the non-stationary Navier-Stokes equation an implicit time discretization method is used and the nonlinearity is linearized via a fixed point strategy in which in the nonlinear term $(\mathbf{u}\cdot\nabla)\mathbf{u}$ is linearized by replacing the first $\mathbf{u}$ argument by an already computed approximation $\mathbf{u}^{\text{old}} =: \mathbf{w}$. For the case $\mathbf{w} = 0$ this problem reduces to the so-called *generalized Stokes equations* and for $\mathbf{w} = 0$ and $\xi = 0$ we obtain the *stationary Stokes problem*.

Formulation in dimensionless variables

For the derivation and analysis of numerical methods for the models presented above it is convenient to consider these models in a non-dimensionalized form. For this we introduce

$$\begin{aligned} L &: \text{ typical length scale (related to size of } \Omega), \\ U &: \text{ typical velocity size,} \end{aligned}$$

and dimensionless variables

$$\bar{x} = \frac{1}{L}x, \quad \bar{t} = \frac{U}{L}t, \quad \bar{\mathbf{u}}(\bar{x},\bar{t}) = \frac{\mathbf{u}(x,t)}{U}, \quad \bar{p}(\bar{x},\bar{t}) = \frac{p(x,t)}{\rho U^2}.$$

Furthermore, $\bar{\Omega} := \frac{1}{L}\Omega := \big\{\, \bar{x} \in \mathbb{R}^3 : \ L\bar{x} \in \Omega \,\big\}$, and a non-dimensional source term is defined as $\bar{\mathbf{g}} = \frac{L}{U^2}\mathbf{g}$. The partial differential equations in (1.14) can be written in these dimensionless quantities as follows, where differential operators w.r.t. $\bar{x}_i$ and $\bar{t}$ are denoted with a $^-$ (for example: $\bar{\nabla}$):

$$\begin{aligned} \frac{\partial \bar{\mathbf{u}}}{\partial \bar{t}} + (\bar{\mathbf{u}}\cdot\bar{\nabla})\bar{\mathbf{u}} + \bar{\nabla}\bar{p} - \frac{1}{Re}\bar{\Delta}\bar{\mathbf{u}} &= \bar{\mathbf{g}} \text{ in } \ \bar{\Omega} \\ \overline{\operatorname{div}}\bar{\mathbf{u}} &= 0 \text{ in } \ \bar{\Omega}, \end{aligned}$$

with the dimensionless *Reynolds number*

$$Re = \frac{\rho L U}{\mu}.$$

For notational simplicity, in the remainder we drop the bar notation, and thus obtain the following Navier-Stokes system in dimensionless variables:

Navier-Stokes

$$\begin{aligned} \frac{\partial \mathbf{u}}{\partial t} - \frac{1}{Re}\Delta \mathbf{u} + (\mathbf{u}\cdot\nabla)\mathbf{u} + \nabla p &= \mathbf{g} \text{ in } \Omega \\ \operatorname{div} \mathbf{u} &= 0 \text{ in } \Omega. \end{aligned} \tag{1.18}$$

We now list the above-mentioned related simpler models used in fluid dynamics. In these lecture notes we use these simpler models in the analysis of numerical methods. The non-stationary Stokes equations (1.16) in dimensionless formulation are as follows:

Stokes

$$\begin{aligned} \frac{\partial \mathbf{u}}{\partial t} - \frac{1}{Re}\Delta \mathbf{u} + \nabla p &= \mathbf{g} \text{ in } \Omega \\ \operatorname{div} \mathbf{u} &= 0 \text{ in } \Omega. \end{aligned} \tag{1.19}$$

Hence, the Stokes equations can be seen as a limit case of the Navier-Stokes equations if $Re \to 0$. In the remainder, if we refer to the (Navier-)Stokes problem we always mean the *non-stationary* (Navier-)Stokes equations. We now present three time *in*dependent models. The Oseen system (1.17) in dimensionless form is as follows:

Oseen

$$\begin{aligned} \xi\mathbf{u} - \frac{1}{Re}\Delta \mathbf{u} + (\mathbf{w}\cdot\nabla)\mathbf{u} + \nabla p &= \mathbf{g} \text{ in } \Omega \\ \operatorname{div} \mathbf{u} &= 0 \text{ in } \Omega. \end{aligned} \tag{1.20}$$

where now $\xi \geq 0$ is a dimensionless constant. *Special cases* of the Oseen problem are the generalized Stokes and the stationary Stokes problems:

Generalized Stokes

$$\begin{aligned} \xi\mathbf{u} - \frac{1}{Re}\Delta \mathbf{u} + \nabla p &= \mathbf{g} \text{ in } \Omega \\ \operatorname{div} \mathbf{u} &= 0 \text{ in } \Omega, \end{aligned} \tag{1.21}$$

Stationary Stokes

$$\begin{aligned} -\frac{1}{Re}\Delta \mathbf{u} + \nabla p &= \mathbf{g} \text{ in } \Omega \\ \operatorname{div} \mathbf{u} &= 0 \text{ in } \Omega. \end{aligned} \tag{1.22}$$

1.3 Weak formulations

We will use finite element methods for the discretization of the (Navier-)Stokes equations. These methods are based on the weak (or variational) formulation of the partial differential equations. In this section we discuss this weak formulation.

1.3.1 Function spaces

We only recall some basic facts from the theory on Sobolev spaces. For a detailed treatment of this subject we refer to the literature, e.g. [1].

One main motivation for using Sobolev spaces is that these are Banach spaces. Some of these are Hilbert spaces. In our treatment of elliptic boundary value problems we only need these Hilbert spaces and thus we restrict ourselves to the presentation of this subset of Sobolev Hilbert spaces.

First we introduce the concept of weak derivatives. Let $\Omega \subset \mathbb{R}^d$ be an open, bounded and connected domain, and take $u \in C^1(\Omega)$, $\phi \in C_0^\infty(\Omega)$, where $C_0^\infty(\Omega)$ consists of all functions in $C^\infty(\Omega)$ that have a compact support in Ω (and thus, since Ω is open, such functions are identically zero close to the boundary). Since ϕ vanishes identically outside some compact subset of Ω, one obtains by partial integration in the variable x_j:

$$\int_\Omega \frac{\partial u(x)}{\partial x_j}\phi(x)\,dx = -\int_\Omega u(x)\frac{\partial \phi(x)}{\partial x_j}\,dx$$

and thus

$$\int_\Omega D^\alpha u(x)\phi(x)\,dx = -\int_\Omega u(x)D^\alpha\phi(x)\,dx, \qquad |\alpha| = 1,$$

holds. Here $D^\alpha u$ with $\alpha = (\alpha_1, \ldots, \alpha_d)$, $|\alpha| = \alpha_1 + \ldots + \alpha_d$ denotes

$$D^\alpha u = \frac{\partial^{|\alpha|}}{\partial x_1^{\alpha_1} \ldots \partial x_d^{\alpha_d}}.$$

Repeated application of this result yields the fundamental *Green's formula*

$$\begin{aligned} &\int_\Omega D^\alpha u(x)\phi(x)\,dx = (-1)^{|\alpha|}\int_\Omega u(x)D^\alpha\phi(x)\,dx, \\ &\text{for all } \ \phi \in C_0^\infty(\Omega),\ u \in C^k(\Omega),\ k = 1, 2, \ldots \quad \text{and } \ |\alpha| \le k. \end{aligned} \tag{1.23}$$

Based on this formula we introduce the notion of a weak derivative:

Definition 1.3.1 Consider $u \in L^2(\Omega)$ and $|\alpha| > 0$. If there exists $v \in L^2(\Omega)$ such that

$$\int_\Omega v(x)\phi(x)\,dx = (-1)^{|\alpha|}\int_\Omega u(x)D^\alpha\phi(x)\,dx \quad \text{for all } \ \phi \in C_0^\infty(\Omega) \tag{1.24}$$

then v is called the αth *weak derivative* of u and is denoted by $D^\alpha u := v$.

Such weak derivatives are often introduced in the more general setting of so-called *distributions*. For our purposes, however, the definition above suffices. If for $u \in L^2(\Omega)$ the αth weak derivative exists then it is unique (in the usual Lebesgue sense). If $u \in C^k(\overline{\Omega})$ then for $0 < |\alpha| \le k$ the αth weak derivative and the classical αth derivative coincide. The Sobolev space $H^m(\Omega)$, $m = 1, 2, \ldots$, consists of all functions in $L^2(\Omega)$ for which all αth weak derivatives with $|\alpha| \le m$ exist:

$$H^m(\Omega) := \left\{ u \in L^2(\Omega) :\ D^\alpha u \ \text{ exists for all } \ 0 < |\alpha| \le m \right\}. \tag{1.25}$$

For $m = 0$ we define $H^0(\Omega) := L^2(\Omega)$. In $H^m(\Omega)$ a natural inner product and corresponding norm are defined by

$$(u,v)_m := \sum_{|\alpha|\le m} (D^\alpha u, D^\alpha v)_{L^2}, \qquad \|u\|_m := (u,u)_m^{\frac{1}{2}}, \qquad u,v \in H^m(\Omega). \tag{1.26}$$

It is easy to verify, that $(\cdot,\cdot)_m$ defines an inner product on $H^m(\Omega)$. We now formulate a main result:

Theorem 1.3.2 *The space* $\left(H^m(\Omega)\,,\,(\cdot,\cdot)_m\right)$ *is a Hilbert space.*

Similar constructions can be applied if we replace the Hilbert space $L^2(\Omega)$ by the Banach space $L^p(\Omega)$, $1 \le p < \infty$ of measurable functions for which $\|u\|_p := (\int_\Omega |u(x)|^p\,dx)^{1/p}$ is bounded. This results in Sobolev spaces which are usually denoted by $H_p^m(\Omega)$. For notational simplicity we deleted the index $p = 2$ in our presentation. For $p \neq 2$ the Sobolev space $H_p^m(\Omega)$ is a Banach space but *not* a Hilbert space.

One can also define these Sobolev spaces using a different technique, namely based on the concept of completion. Consider the function space

$$Z_m := \{\, u \in C^\infty(\Omega) :\, \|u\|_m < \infty \,\}.$$

The completion of this space with respect to $\|\cdot\|_m$ yields the Sobolev space $H^m(\Omega)$:

$$H^m(\Omega) = \text{ completion of } \ (Z_m\,,\,(\cdot,\cdot)_m).$$

A compact notation is $H^m(\Omega) = \overline{Z_m}^{\|\cdot\|_m}$. Note that $C^\infty(\overline{\Omega}) \subset Z_m$. Under very mild assumptions on the domain Ω we even have

$$H^m(\Omega) = \overline{C^\infty(\overline{\Omega})}^{\|\cdot\|_m}.$$

Another space that plays an important role in the weak formulation of the partial differential equations is the following subspace of $H^1(\Omega)$:

$$H_0^1(\Omega) := \overline{C_0^\infty(\Omega)}^{\|\cdot\|_1}. \tag{1.27}$$

An important issue is the smoothness in the classical sense of functions from a Sobolev space. In this respect the following Sobolev embedding result is relevant:

$$H^m(\Omega) \hookrightarrow C^k(\overline{\Omega}) \quad \text{if } \ m - \frac{d}{2} > k \quad (d \text{ such that } \Omega \subset \mathbb{R}^d). \tag{1.28}$$

The symbol $\hookrightarrow$ is used to denote that the embedding between the two spaces is continuous, i.e. if m and k satisfy the condition in (1.28) then there exists a constant c such that

$$\|u\|_{C^k(\overline{\Omega})} \le c\|u\|_m \quad \text{for all } \ u \in H^m(\Omega).$$

A basic result is the so-called Poincaré-Friedrichs inequality:

$$\|u\|_{L^2} \le C \sqrt{\sum_{|\alpha|=1} \|D^\alpha u\|_{L^2}^2} \quad \text{for all } \ u \in H_0^1(\Omega). \tag{1.29}$$

Based on this we have the following norm equivalence:

$$|u|_1 \le \|u\|_1 \le C|u|_1 \quad \text{for all } \ u \in H_0^1(\Omega), \ \ |u|_1^2 := \sum_{|\alpha|=1} \|D^\alpha u\|_{L^2}^2, \tag{1.30}$$

i.e., $|\cdot|_1$ and $\|\cdot\|_1$ are *equivalent norms* on $H_0^1(\Omega)$.

In the weak formulation of elliptic boundary value problems one has to treat boundary conditions. For this the next result will be needed.

There exists a unique bounded linear operator

$$\gamma : H^1(\Omega) \to L^2(\partial\Omega), \quad \|\gamma(u)\|_{L^2(\partial\Omega)} \leq c\|u\|_1, \tag{1.31}$$

with the property that for all $u \in C^1(\overline{\Omega})$ the equality $\gamma(u) = u_{|\partial\Omega}$ holds. The operator γ is called the *trace operator*. For $u \in H^1(\Omega)$ the function $\gamma(u) \in L^2(\partial\Omega)$ represents the boundary "values" of u and is called the trace of u. For $\gamma(u)$ one often uses the notation $u_{|\partial\Omega}$. For example, for $u \in H^1(\Omega)$, the identity $u_{|\partial\Omega} = 0$ means that $\gamma(u) = 0$ in the $L^2(\partial\Omega)$ sense. Using the trace operator one can give another natural characterization of the space $H_0^1(\Omega)$:

$$H_0^1(\Omega) = \left\{ u \in H^1(\Omega) : \, u_{|\partial\Omega} = 0 \right\}.$$

The *dual space* of $H_0^1(\Omega)$, i.e. the space of all bounded linear functionals $H_0^1(\Omega) \to \mathbb{R}$ is denoted by

$$H^{-1}(\Omega) := H_0^1(\Omega)',$$

with norm

$$\|f\|_{-1} := \sup_{v \in H_0^1(\Omega)} \frac{f(v)}{\|v\|_1}, \quad f \in H^{-1}(\Omega).$$

Finally, we collect a few results on *Green's formulas* that hold in Sobolev spaces. For notational simplicity the function arguments x are deleted in the integrals, and in boundary integrals like, for example, $\int_{\partial\Omega} \gamma(u)\,\gamma(v)\,ds$ we delete the trace operator γ. The following identities hold, with $\mathbf{n} = (n_1, \ldots, n_d)$ the outward unit normal on $\partial\Omega$ and $H^m := H^m(\Omega)$:

$$\begin{aligned}
\int_\Omega u \frac{\partial v}{\partial x_i}\, dx &= -\int_\Omega \frac{\partial u}{\partial x_i} v\, dx + \int_{\partial\Omega} uvn_i\, ds, \quad u, v \in H^1, 1 \leq i \leq d \\
\int_\Omega \Delta u\, v\, dx &= -\int_\Omega \nabla u \cdot \nabla v\, dx + \int_{\partial\Omega} \nabla u \cdot \mathbf{n} v\, ds, \quad u \in H^2, v \in H^1 \\
\int_\Omega u\, \operatorname{div} \mathbf{v}\, dx &= -\int_\Omega \nabla u \cdot \mathbf{v}\, dx + \int_{\partial\Omega} u\mathbf{v} \cdot \mathbf{n}\, ds, \quad u \in H^1, \mathbf{v} \in (H^1)^d.
\end{aligned}$$

The Sobolev spaces turn out to be the appropriate ones for the weak formulation of many partial differential equations. For the weak formulation of *time* dependent partial differential equations one needs in addition the concept of V-valued functions, where V is a given Banach space (for example $L^2(\Omega)$, or a Sobolev space). We now introduce this concept. For proofs and more information we refer to the literature, e.g. [45, 46].

Let V be a Banach space with norm denoted by $\|\cdot\|_V$. The space $L^2(0,T;V)$ consists of all functions $u : \, (0,T) \to V$ for which

$$\|u\|_{L^2(0,T;V)} := \Big(\int_0^T \|u(t)\|_V^2\, dt \Big)^{\frac{1}{2}} < \infty$$

holds. The space $L^2(0,T;V)$ is a Banach space. It is a Hilbert space if V is a Hilbert space. This definition applied to the dual space V' results in the space $L^2(0,T;V')$ of functional valued functions $u : \, (0,T) \to V'$ with

$$\|u\|_{L^2(0,T;V')} := \Big(\int_0^T \|u(t)\|_{V'}^2\, dt \Big)^{\frac{1}{2}} < \infty.$$

Recall that $\|u(t)\|_{V'} := \sup_{v \in V} \frac{|u(t)(v)|}{\|v\|_V}$. There is a linear bijective isometric mapping $j : \, L^2(0,T;V)' \to L^2(0,T;V')$, hence these spaces can be identified with each other. One often writes

$$L^2(0,T;V)' = L^2(0,T;V').$$

1.3.2 Oseen problem in weak formulation

In this section we treat the weak formulation of the Oseen problem in dimensionless form (1.20). For notational simplicity we only treat the three-dimensional case, i.e. $\Omega \subset \mathbb{R}^3$. We consider this problem with homogeneous Dirichlet boundary conditions $\mathbf{u} = 0$ on $\partial\Omega$. The Reynolds number $Re > 0$ and the problem parameter $\xi \geq 0$ are given constants. Furthermore we assume that the velocity field $\mathbf{w}$ satisfies $\mathbf{w} \in H^1(\Omega)^3$ and $\|\mathbf{w}\|_{L^\infty(\Omega)} < \infty$. We introduce the spaces

$$\mathbf{V} := H_0^1(\Omega)^3, \quad Q := L_0^2(\Omega) = \left\{ q \in L^2(\Omega) : \int_\Omega q \, dx = 0 \right\}, \tag{1.32}$$

and the bilinear forms

$$m(\mathbf{u}, \mathbf{v}) = (\mathbf{u}, \mathbf{v})_{L^2} = \int_\Omega \mathbf{u} \cdot \mathbf{v} \, dx \tag{1.33a}$$

$$a(\mathbf{u}, \mathbf{v}) = \frac{1}{Re} \int_\Omega \nabla \mathbf{u} \cdot \nabla \mathbf{v} \, dx = \frac{1}{Re} \sum_{i=1}^{3} \int_\Omega \nabla u_i \cdot \nabla v_i \, dx \tag{1.33b}$$

$$c(\mathbf{u}, \mathbf{v}) = c(\mathbf{w}; \mathbf{u}, \mathbf{v}) = \int_\Omega (\mathbf{w} \cdot \nabla \mathbf{u}) \cdot \mathbf{v} \, dx = \sum_{1 \leq i,j \leq 3} \int_\Omega w_i \frac{\partial u_j}{\partial x_i} v_j \, dx \tag{1.33c}$$

$$b(\mathbf{v}, q) = -\int_\Omega q \operatorname{div} \mathbf{v} \, dx. \tag{1.33d}$$

The weak formulation of (1.20) is as follows:

Find $\mathbf{u} \in \mathbf{V}$ and $p \in Q$ such that

$$\begin{aligned} \xi m(\mathbf{u}, \mathbf{v}) + a(\mathbf{u}, \mathbf{v}) + c(\mathbf{u}, \mathbf{v}) + b(\mathbf{v}, p) &= (\mathbf{g}, \mathbf{v})_{L^2} \quad \text{for all} \quad \mathbf{v} \in \mathbf{V}, \\ b(\mathbf{u}, q) &= 0 \quad \text{for all} \quad q \in Q. \end{aligned} \tag{1.34}$$

We now turn to the analysis of this weak formulation. First we show that if the problem in strong formulation (1.20) has a sufficiently smooth solution $(\mathbf{u}, p)$, then this pair is also a solution of (1.34). We start with the observation that the bilinear forms $m(\cdot,\cdot)$, $a(\cdot,\cdot)$, $c(\cdot,\cdot)$ are *continuous* on $\mathbf{V} \times \mathbf{V}$ and that $b(\cdot,\cdot)$ is *continuous* on $\mathbf{V} \times Q$.

Lemma 1.3.3 *Assume $\mathbf{u} \in C^2(\overline{\Omega})^3$ with $\mathbf{u} = 0$ on $\partial\Omega$ and $p \in C^1(\overline{\Omega})$ with $\int_\Omega p \, dx = 0$ is a solution pair of* (1.20)*. Then this pair also solves* (1.34)*.*

Proof. From the assumptions on $\mathbf{u}$ and p it follows that $\mathbf{u} \in \mathbf{V}$, $p \in Q$. If we multiply the first equation in (1.20) by $\boldsymbol{\phi} \in C_0^\infty(\Omega)^3$, integrate over Ω and apply partial integration (for each of the three components) we obtain

$$\xi \int_\Omega \mathbf{u}\boldsymbol{\phi} \, dx + \frac{1}{Re} \int_\Omega \nabla \mathbf{u} \cdot \nabla \boldsymbol{\phi} \, dx + \int_\Omega (\mathbf{w} \cdot \nabla \mathbf{u}) \cdot \boldsymbol{\phi} \, dx - \int_\Omega p \operatorname{div} \boldsymbol{\phi} \, dx = \int_\Omega \mathbf{g} \cdot \boldsymbol{\phi} \, dx.$$

Hence,

$$\xi m(\mathbf{u}, \boldsymbol{\phi}) + a(\mathbf{u}, \boldsymbol{\phi}) + c(\mathbf{u}, \boldsymbol{\phi}) + b(\boldsymbol{\phi}, p) = (\mathbf{g}, \boldsymbol{\phi})_{L^2} \tag{1.35}$$

for all $\boldsymbol{\phi} \in C_0^\infty(\Omega)^3$. Using the continuity of the bilinear forms and of $\mathbf{v} \to (\mathbf{g}, \mathbf{v})_{L^2}$ on $\mathbf{V}$, and the density of $C_0^\infty(\Omega)^3$ in $\mathbf{V}$ it follows that the identity in (1.35) even holds for all $\boldsymbol{\phi} \in \mathbf{V}$. Thus the first variational equation in (1.34) holds. Multiplication of the second equation in (1.20) by an arbitrary $q \in Q$ and integrating over Ω results in the second variational identity in (1.34). □

One very important property of the weak formulation (1.34) is that, opposite to the strong formulation in (1.34), under very mild assumptions it has a unique solution. The mathematical analysis of variational problems like the one in (1.34) is based on an abstract theory for saddle point problems as presented in the Appendix, Section 4.3. Theorem 4.3.1 can be applied to prove the well-posedness of the weak formulation of the Oseen problem (1.34). In the application of Theorem 4.3.1 we use the spaces $V = \mathbf{V},\ M = Q$ and the bilinear forms

$$\hat{a}(\mathbf{u},\mathbf{v}) = \xi m(\mathbf{u},\mathbf{v}) + a(\mathbf{u},\mathbf{v}) + c(\mathbf{u},\mathbf{v}) \quad \text{on} \quad \mathbf{V}\times\mathbf{V}, \tag{1.36a}$$

$$\hat{b}(\mathbf{u},q) = b(\mathbf{u},q) \quad \text{on} \quad \mathbf{V}\times Q. \tag{1.36b}$$

A rather deep result from the theory on Sobolev spaces is the following:

$$\exists\ \beta>0\ :\quad \sup_{\mathbf{v}\in H_0^1(\Omega)^d} \frac{\int_\Omega q\ \mathrm{div}\,\mathbf{v}\,dx}{\|\mathbf{v}\|_1} \ \geq\ \beta\,\|q\|_{L^2} \quad \forall\ q\in L_0^2(\Omega). \tag{1.37}$$

A proof of this is given in [33, 15]. From this result we immediately obtain that for the bilinear form $b(\cdot,\cdot)$ on $\mathbf{V}\times Q$ the inf-sup condition in (4.17a) is satisfied. Using this the following theorem can be proved:

Theorem 1.3.4 *Consider the weak formulation of the Oseen problem in* (1.34)*. Assume that* ξ *and* $\mathbf{w}$ *are such that* $\xi - \frac{1}{2}\mathrm{div}\,\mathbf{w} \geq 0$ *on* Ω*. Then this problem is well-posed.*

Proof. We apply Theorem 4.3.1 with the spaces $V = \mathbf{V}$, $M = Q$, the bilinear forms defined in (1.36) and the functionals $f_1(\mathbf{v}) := (\mathbf{g},\mathbf{v})$, $f_2 = 0$. These bilinear forms and functionals are continuous. The inf-sup condition (4.17a) is satisfied due to property (1.37). We finally check the ellipticity condition (4.17b). For $\mathbf{u},\mathbf{v}\in\mathbf{V}$ we have $\mathbf{u}_{|\partial\Omega} = \mathbf{v}_{|\partial\Omega} = 0$ and thus using partial integration we obtain

$$\begin{aligned}\int_\Omega (\mathbf{w}\cdot\nabla\mathbf{u})\cdot\mathbf{v}\,dx &= \sum_{1\leq i,j\leq 3}\int_\Omega w_i\frac{\partial u_j}{\partial x_i}v_j\,dx \\ &= -\int_\Omega \mathrm{div}\,\mathbf{w}(\mathbf{u}\cdot\mathbf{v})\,dx - \int_\Omega (\mathbf{w}\cdot\nabla\mathbf{v})\cdot\mathbf{u}\,dx.\end{aligned}$$

Hence, for $\mathbf{u}=\mathbf{v}$ we have

$$\int_\Omega (\mathbf{w}\cdot\nabla\mathbf{u})\cdot\mathbf{u}\,dx = -\frac{1}{2}\int_\Omega \mathrm{div}\,\mathbf{w}(\mathbf{u}\cdot\mathbf{u})\,dx.$$

This yields, using the assumption $\xi - \frac{1}{2}\mathrm{div}\,\mathbf{w}\geq 0$,

$$\begin{aligned}\hat{a}(\mathbf{u},\mathbf{u}) &= \xi\int_\Omega \mathbf{u}\cdot\mathbf{u}\,dx + \frac{1}{Re}\int_\Omega \nabla\mathbf{u}\cdot\nabla\mathbf{u}\,dx + \int_\Omega (\mathbf{w}\cdot\nabla\mathbf{u})\cdot\mathbf{u}\,dx \\ &= \int_\Omega \big(\xi - \frac{1}{2}\mathrm{div}\,\mathbf{w}\big)\mathbf{u}\cdot\mathbf{u}\,dx + \frac{1}{Re}\int_\Omega \nabla\mathbf{u}\cdot\nabla\mathbf{u}\,dx \\ &\geq \frac{1}{Re}\int_\Omega \nabla\mathbf{u}\cdot\nabla\mathbf{u}\,dx \geq c\,\|\mathbf{u}\|_1^2,\end{aligned}$$

with a constant $c>0$. In the last inequality we used the norm equivalence (1.30). □

We comment on the assumption $\xi - \frac{1}{2}\mathrm{div}\,\mathbf{w}\geq 0$ on Ω in Theorem 1.3.4. If we consider a stationary Stokes problem then we have $\xi = 0$, $\mathbf{w} = 0$ and thus this assumption is satisfied. For a generalized Stokes equation, which results from implicit time discretization of an non-stationary Stokes problem we have $\mathbf{w} = 0$, $\xi > 0$ and thus the assumption holds. Hence we conclude:

The stationary (generalized) Stokes equations in weak formulation, i.e. (1.34) *with* $\xi \geq 0$, $\mathbf{w} = 0$, *have a unique solution pair* $(\mathbf{u}, p) \in \mathbf{V} \times Q$.

In the general case of an Oseen equation, which results after implicit time integration and linearization of a Navier-Stokes problem, it is reasonable to expect that $|\operatorname{div} \mathbf{w}|$, in which $\mathbf{w}$ is an approximation of the solution $\mathbf{u}$ (that satisfies $\operatorname{div} \mathbf{u} = 0$), is small compared to $\xi \sim \frac{1}{\Delta t}$ (Δt: time step in time discretization). Hence it is plausible that the condition $\xi - \frac{1}{2} \operatorname{div} \mathbf{w} \geq 0$ is satisfied.

1.3.3 Time dependent (Navier-)Stokes equations in weak formulation

In this section we treat the weak formulation of the non-stationary Stokes problem (1.19) and of the non-stationary Navier-Stokes problem (1.18). For both problems we restrict ourselves to the case with homogeneous Dirichlet boundary conditions, i.e., $\mathbf{u} = 0$ on $\partial\Omega$.

Compared to the weak formulation of the Oseen problem in Section 1.3.2 we now in addition have to address the issue of an appropriate treatment of the time derivative $\frac{\partial \mathbf{u}}{\partial t}$. For the weak formulation of many time dependent partial differential equations the Hilbert space $L^2(0,T;V)$ with a suitable Sobolev space V, cf. Section 1.3.1, turns out to be appropriate. For $u : (0,T) \to V$ one then needs a suitable weak derivative $u' = \frac{du}{dt}$. This can be defined by means of the very general and powerful concept of distributional derivatives, in which derivatives of linear mappings $L : C_0^\infty(0,T) \to V$ are defined, cf. [45, 46]. We will not use this (rather abstract) approach but introduce u' by means of weak derivatives as already presented in Definition 1.3.1. This (compared to the distributional concept) less general definition is sufficient for our purposes. We first present a weak variational formulation of a time dependent problem in an abstract setting and then apply this to derive weak formulations of the non-stationary Stokes- and Navier-Stokes equations.

An abstract variational formulation of a time dependent problem

Let V, H be Hilbert spaces such that $V \hookrightarrow H \hookrightarrow V'$ forms a Gelfand triple, which means that H is identified with its dual, $H \equiv H'$, the embedding $V \hookrightarrow H$ is continuous and V is dense in H. The scalar products in V and H are denoted by $(\cdot,\cdot)_V$ and $(\cdot,\cdot)_H$, respectively. In our applications we use, for example, the Gelfand triple $H_0^1(\Omega) \hookrightarrow L^2(\Omega) \hookrightarrow H^{-1}(\Omega)$.

We recall the definition of a weak derivative for a function $g \in L^2(0,T)$ as given in Definition 1.3.1: if for such a function g there exists $h \in L^2(0,T)$ such that

$$\int_0^T h(t)\phi(t)\,dt = -\int_0^T g(t)\phi'(t)\,dt \quad \text{for all} \quad \phi \in C_0^\infty(0,T), \tag{1.38}$$

then $h =: g'$ is the weak derivative of g.

We now introduce a weak time derivative for $u \in L^2(0,T;V)$. It is possible to define a weak derivative of u in the same space $L^2(0,T;V)$. However, for the time dependent problems that we consider it turns out to be more appropriate to define a weak derivative in the space $L^2(0,T;V')$. Using the Gelfand property $V \hookrightarrow H \equiv H' \hookrightarrow V'$ this can be done as follows. Take $t \in [0,T]$. The element $u(t) \in V$ can be identified with an element of V', through $v \to (u(t), v)_H$, $v \in V$. This identification of $u(t)$ with an element of V' leads to the following natural definition of a weak time derivative of u in $L^2(0,T;V')$.

Definition 1.3.5 Consider $u \in L^2(0,T;V)$. If there exists a $w \in L^2(0,T;V')$ such that

$$\int_0^T w(t)(v)\,\phi(t)\,dt = -\int_0^T (u(t),v)_H \phi'(t)\,dt, \tag{1.39}$$

for all $v \in V$ and all $\phi \in C_0^\infty(0,T)$, then w is called the weak (time) derivative of u and we write $u' = w$.

One can show, that if such a weak derivative exists, then it is unique. Furthermore, assume that $u : [0,T] \to H$ is smooth enough such that the classical (Fréchet) derivative exists in H. Denote this Fréchet derivative by $u'(t)$. Then

$$\int_0^T (u'(t), v)_H \, \phi(t) \, dt = - \int_0^T (u(t), v)_H \phi'(t) \, dt,$$

holds for all $v \in V$ and all $\phi \in C_0^\infty(0,T)$, and thus $u'(t)$ identified with the functional $v \to (u'(t), v)_H$ is the weak derivative in the sense of Definition 1.3.5.

Lemma 1.3.6 *Assume that $u \in L^2(0,T;V)$ has a weak derivative $u' \in L^2(0,T;V')$. For arbitrary $v \in V$ define the function $g_v : t \to (u(t), v)_H$. Then $g_v \in L^2(0,T)$ and g_v has a weak derivative $g_v'(t) = \frac{d}{dt}(u(t), v)_H$ in the sense of* (1.38)*. Furthermore,*

$$\frac{d}{dt}(u(t), v)_H = u'(t)(v) \quad \textit{for all} \;\; v \in V \tag{1.40}$$

holds for almost all $t \in (0,T)$.

Proof. Take $v \in V$. Note that due to the Gelfand property $\|w\|_H \le c\|w\|_V$ for all $w \in V$ holds. From

$$\int_0^T g_v(t)^2 \, dt \le \int_0^T \|u(t)\|_H^2 \|v\|_H^2 \, dt \le c^4 \int_0^T \|u(t)\|_V^2 \, dt \, \|v\|_V^2$$

and $u \in L^2(0,T;V)$ it follows that $g_v \in L^2(0,T)$. For $h_v(t) := u'(t)(v)$ we have

$$\int_0^T h_v(t)^2 \, dt \le \int_0^T \|u'(t)\|_{V'}^2 \, dt \|v\|_V^2 < \infty$$

and thus $h_v \in L^2(0,T)$. Using the property (1.39) we get

$$\begin{aligned} \int_0^T h_v(t)\phi(t) \, dt &= \int_0^T u'(t)(v)\phi(t) \, dt \\ &= - \int_0^T (u(t), v)_H \phi'(t) \, dt = - \int_0^T g_v(t)\phi'(t) \, dt \end{aligned}$$

for all $\phi \in C_0^\infty(0,T)$. Thus $h_v = g_v'$, i.e., $u'(t)(v) = \frac{d}{dt}(u(t), v)_H$ holds. □

Using the above notion of a weak derivative for functions $u \in L^2(0,T;V)$ we introduce the following space

$$W^1(0,T;V) := \left\{ v \in L^2(0,T;V) : \, v' \in L^2(0,T;V') \;\; \text{exists} \right\}.$$

We mention two important properties of this space. For proofs and further properties we refer to the literature, e.g. [45, 46]. Firstly, the space $W^1(0,T;V)$ is a Hilbert space w.r.t. the norm (and corresponding scalar product)

$$\|u\|_{W^1(0,T;V)} := \left(\|u\|_{L^2(0,T;V)}^2 + \|u'\|_{L^2(0,T;V')}^2 \right)^{\frac{1}{2}}.$$

Secondly, there is a continuous embedding

$$W^1(0,T;V) \hookrightarrow C([0,T];H), \tag{1.41}$$

where $C([0,T];H)$ is the Banach space of all continuous functions $u : [0,T] \to H$ with norm $\|u\|_{C([0,T];H)} := \max_{0 \le t \le T} \|u(t)\|_H$. An important corollary of this embedding property is that for $u \in W^1(0,T;V)$ the function $u(0) := \lim_{t \downarrow 0} u(t)$ is well-defined in H.

Based on these preparations, we can introduce an abstract variational time dependent problem. Let $\hat{a} : V \times V \to \mathbb{R}$ be a bilinear form. For $b(t) \in V'$, $t \in (0,T)$, $u_0 \in H$ we define the problem:

Find $u \in W^1(0,T;V)$ such that

$$\begin{aligned} \frac{d}{dt}(u(t),v)_H + \hat{a}(u(t),v) &= b(t)(v) \quad \text{for all } \ v \in V,\ t \in (0,T), \\ u(0) &= u_0. \end{aligned} \tag{1.42}$$

Due to Lemma 1.3.6 the term $\frac{d}{dt}(u(t),v)_H$ in (1.42) is well-defined. A main theorem is the following, cf. [45, 46] for a proof.

Theorem 1.3.7 *Take $u_0 \in H$, $b \in L^2(0,T;V')$ and assume that the bilinear form $\hat{a}(\cdot,\cdot)$ is continuous and elliptic on $V \times V$. Then the variational problem (1.42) is well-posed, i.e. it has a unique solution u and the linear mapping $(b,u_0) \to u$ is continuous from $L^2(0,T;V') \times H$ into $W^1(0,T;V)$.*

We summarize the essential ingredients for this abstract time dependent weak formulation to be well-posed:

- One uses a Gelfand triple of spaces $V \hookrightarrow H \hookrightarrow V'$ and uses the corresponding space $W^1(0,T;V)$. The weak derivative $u' \in L^2(0,T;V')$ is as in Definition 1.3.5.
- The data are from appropriate spaces: $b \in L^2(0,T;V')$, $u_0 \in H$.
- The bilinear form $\hat{a}(\cdot,\cdot)$ is continuous and elliptic on $V \times V$.

Below we use these abstract results for the derivation of appropriate weak formulations of the time dependent (Navier-)Stokes equations.

Remark 1.3.8 A proof of Theorem 1.3.7 is given in e.g. Theorem 26.1 in [45], Theorem 23.A in [46]. There it is also shown that the result still holds if the ellipticity condition $\hat{a}(v,v) \geq \gamma_V \|v\|_V^2$ for all $v \in V$ ($\gamma_V > 0$) is replaced by the weaker so-called Garding inequality:

$$\hat{a}(v,v) \geq \gamma_V \|v\|_V^2 - \gamma_H \|v\|_H^2 \quad \text{for all } \ v \in V,$$

with constants $\gamma_V > 0$ and γ_H independent of v.

Application to non-stationary Stokes equations

First we introduce suitable function spaces. Let

$$N(\Omega) := \{\, \mathbf{v} \in C_0^\infty(\Omega)^3 \mid \operatorname{div} \mathbf{v} = 0 \,\}$$

and

$$\mathbf{H}_{\operatorname{div}} := \overline{N(\Omega)}^{\|\cdot\|_{L^2}} \quad \text{(closure of } N(\Omega) \text{ in } L^2(\Omega)^3 \text{)} \tag{1.43}$$

$$\mathbf{V}_{\operatorname{div}} := \overline{N(\Omega)}^{\|\cdot\|_1} \quad \text{(closure of } N(\Omega) \text{ in } H_0^1(\Omega)^3 \text{)}. \tag{1.44}$$

The spaces $\big(\mathbf{H}_{\operatorname{div}}, \|\cdot\|_{L^2}\big)$, $\big(\mathbf{V}_{\operatorname{div}}, \|\cdot\|_1\big)$ are Hilbert spaces. From $\|\mathbf{v}\|_{L^2} \leq c\|\mathbf{v}\|_1$ for all $\mathbf{v} \in N(\Omega)$ and a density argument it follows that there is a continuous embedding $\mathbf{V}_{\operatorname{div}} \hookrightarrow \mathbf{H}_{\operatorname{div}}$. Using $N(\Omega) \subset \mathbf{V}_{\operatorname{div}} \subset \mathbf{H}_{\operatorname{div}}$ and the fact that $\mathbf{H}_{\operatorname{div}}$ is the closure of $N(\Omega)$ w.r.t. $\|\cdot\|_{L^2}$ it follows that $\mathbf{V}_{\operatorname{div}}$ is dense in $\mathbf{H}_{\operatorname{div}}$. Thus we have a Gelfand triple

$$\mathbf{V}_{\operatorname{div}} \hookrightarrow \mathbf{H}_{\operatorname{div}} \equiv \mathbf{H}_{\operatorname{div}}' \hookrightarrow \mathbf{V}_{\operatorname{div}}'. \tag{1.45}$$

The space $\mathbf{V}_{\operatorname{div}}$ can also be characterized as, cf. [39],

$$\mathbf{V}_{\operatorname{div}} = \left\{\, \mathbf{v} \in H_0^1(\Omega)^3 : \operatorname{div} \mathbf{v} = 0 \,\right\}.$$

We take the bilinear form $a(\cdot,\cdot)$ as in (1.33b). It is continuous and elliptic on $H_0^1(\Omega)^3$, thus also on the subspace $\mathbf{V}_{\operatorname{div}}$ of $H_0^1(\Omega)^3$.

We introduce the following weak formulation of the non-stationary Stokes equations (1.19):

Determine $\mathbf{u} \in W^1(0,T;\mathbf{V}_{\mathrm{div}})$ such that

$$\begin{aligned} \frac{d}{dt}(\mathbf{u}(t),\mathbf{v})_{L^2} + a(\mathbf{u}(t),\mathbf{v}) &= (\mathbf{g},\mathbf{v})_{L^2} \quad \text{for all } \ \mathbf{v} \in \mathbf{V}_{\mathrm{div}}, \\ \mathbf{u}(0) &= \mathbf{u}_0 \end{aligned} \tag{1.46}$$

We first show that a smooth solution $(\mathbf{u},p)$ of (1.19) is also a solution of this weak formulation:

Lemma 1.3.9 *Let $(\mathbf{u},p)$ be a solution of (1.19). Define $\mathbf{u}(t) := \mathbf{u}(\cdot,t)$ and assume that $\mathbf{u} \in C^1\big([0,T];C^2(\overline{\Omega})^3\big)$, and $\mathbf{u}(0) = \mathbf{u}_0$. Then $\mathbf{u}$ satisfies (1.46)*

Proof. From $\mathbf{u} \in C^1\big([0,T];C^2(\overline{\Omega})^3\big)$ and $\operatorname{div}\mathbf{u} = 0$ it follows that $\mathbf{u} \in W^1(0,T;\mathbf{V}_{\mathrm{div}})$. We multiply the first equation in (1.19) by $\mathbf{v} \in \mathbf{V}_{\mathrm{div}}$ and integrate over Ω. Note that $\int_\Omega \nabla p \cdot \mathbf{v}\,dx = -\int_\Omega p \operatorname{div}\mathbf{v}\,dx = 0$, due to $\mathbf{v} \in \mathbf{V}_{\mathrm{div}}$. Using partial integration for the term $\frac{1}{Re}\int_\Omega \Delta\mathbf{u}\cdot\mathbf{v}\,dx$ we then obtain

$$(\frac{d\mathbf{u}(t)}{dt},\mathbf{v})_{L^2} + a(\mathbf{u}(t),\mathbf{v}) = (\mathbf{g},\mathbf{v})_{L^2} \quad \text{for all } \ \mathbf{v} \in \mathbf{V}_{\mathrm{div}},$$

and thus (1.46) holds. □

As in Section 1.3.2, opposite to the strong formulation, the weak formulation has the nice property that well-posedness can be shown to hold under (very) mild assumptions on the data. To be more precise, the following theorem holds:

Theorem 1.3.10 *Assume $\mathbf{g} \in L^2(0,T;\mathbf{V}'_{\mathrm{div}})$ and $\mathbf{u}_0 \in \mathbf{H}_{\mathrm{div}}$. Then the weak formulation (1.46) is well-posed.*

Proof. We have a Gelfand triple $\mathbf{V}_{\mathrm{div}} \hookrightarrow \mathbf{H}_{\mathrm{div}} \equiv \mathbf{H}'_{\mathrm{div}} \hookrightarrow \mathbf{V}'_{\mathrm{div}}$ and the bilinear form $a(\cdot,\cdot)$ is continuous and elliptic on $\mathbf{V}_{\mathrm{div}}$. Application of Theorem 1.3.7 yields the desired result. □

The weak formulation in (1.46) has *no pressure variable*. This is due to the fact that the space $\mathbf{V}_{\mathrm{div}}$ contains only functions that satisfy the incompressibility condition $\operatorname{div}\mathbf{u} = 0$. For numerical purposes this weak formulation is less convenient due to the fact that in general it is hard to construct appropriate finite element subspaces of $\mathbf{V}_{\mathrm{div}}$. It turns out to be more convenient to have a weak formulation using $\mathbf{V} = H_0^1(\Omega)^3$ (instead of $\mathbf{V}_{\mathrm{div}}$). This can be achieved by introducing a suitable Lagrange-multiplier variable. In an abstract Hilbert space setting this is explained in Section 4.3, cf. Theorem 4.3.4. As in the strong formulation in (1.19), one can enforce the incompressibility condition by using a pressure variable $p \in Q = L_0^2(\Omega)$. The weak formulation in (1.46) can be used to derive a suitable weak formulation for $(\mathbf{u},p) \in \mathbf{V} \times Q$, which is as follows:

Determine $\mathbf{u} \in W^1(0,T;\mathbf{V}),\ p \in L^2(0,T;Q)$ such that

$$\begin{aligned} \frac{d}{dt}(\mathbf{u}(t),\mathbf{v})_{L^2} + a(\mathbf{u}(t),\mathbf{v}) + b(\mathbf{v},p(t)) &= (\mathbf{g},\mathbf{v})_{L^2} \quad \text{for all } \ \mathbf{v} \in \mathbf{V}, \\ b(\mathbf{u}(t),q) &= 0 \quad \text{for all } \ q \in Q, \\ \mathbf{u}(0) &= \mathbf{u}_0. \end{aligned} \tag{1.47}$$

Theorem 1.3.11 *For $\mathbf{g} \in L^2(0,T;L^2(\Omega)^3)$ and $\mathbf{u}_0 \in \mathbf{V}_{\mathrm{div}}$ the weak formulation (1.47) has a unique solution $(\mathbf{u},p)$. The solution $\mathbf{u}$ also solves the problem (1.46).*

Proof. A complete proof is given in [16] Proposition 6.38. We outline the main ideas of that proof. Assume that (1.47) has a solution $(\mathbf{u}, p) \in W^1(0,T;\mathbf{V}) \times L^2(0,T;Q)$. From the second equation in (1.47) we obtain $\operatorname{div} \mathbf{u} = 0$ in $L^2(\Omega)$. From this and from $\mathbf{u} \in L^2(0,T;\mathbf{V})$ it follows that $\mathbf{u} \in L^2(0,T;\mathbf{V}_{\operatorname{div}})$ holds. Note that $\mathbf{V}' \subset \mathbf{V}'_{\operatorname{div}}$ and thus from $\mathbf{u}' \in L^2(0,T;\mathbf{V}')$ it follows that $\mathbf{u}' \in L^2(0,T;\mathbf{V}'_{\operatorname{div}})$. Hence we have $\mathbf{u} \in W^1(0,T;\mathbf{V}_{\operatorname{div}})$. From the first equation in (1.47) with arbitrary $\mathbf{v} \in \mathbf{V}_{\operatorname{div}}$ (thus $b(\mathbf{v}, p(t)) = 0$) it follows that $\frac{d}{dt}(\mathbf{u}(t), \mathbf{v})_{L^2} + a(\mathbf{u}(t), \mathbf{v}) = (\mathbf{g}, \mathbf{v})_{L^2}$ for all $\mathbf{v} \in \mathbf{V}_{\operatorname{div}}$. We conclude that $\mathbf{u}$ is a solution of (1.46). Furthermore, due to Theorem 1.3.10, we have *uniqueness* of the solution $(\mathbf{u}, p)$. We now address existence of a solution $(\mathbf{u}, p)$ of (1.47). Let $\mathbf{u}$ be the unique solution of (1.46), which exists due to Theorem 1.3.10. From $\operatorname{div} \mathbf{u} = 0$ is follows that $\mathbf{u}$ satisfies the second equation in (1.47). For the solution $\mathbf{u}$ we have $\mathbf{u}' \in L^2(0,T;\mathbf{V}'_{\operatorname{div}})$. One can show (using the assumptions on the data, cf. [16]), that $\mathbf{u}' \in L^2(0,T;\mathbf{V}')$ holds. Hence, $\mathbf{u} \in W^1(0,T;\mathbf{V})$ and moreover,

$$\ell(t)(\mathbf{v}) := \frac{d}{dt}(\mathbf{u}(t), \mathbf{v})_{L^2} + a(\mathbf{u}(t), \mathbf{v}) - (\mathbf{g}, \mathbf{v})_{L^2}$$

satisfies

$$\ell \in L^2(0,T;\mathbf{V}'), \quad \ell(t)(\mathbf{v}) = 0 \quad \text{for all} \;\; t \in [0,T], \;\; \mathbf{v} \in \mathbf{V}_{\operatorname{div}}.$$

From a rather deep result (originally due to De Rham [14]), cf. Theorem 2.3 in [20], it follows that there exists $p(t) \in L^2_0(\Omega) = Q$ such that

$$(p(t), \operatorname{div} \mathbf{v})_{L^2} = \ell(t)(\mathbf{v}) \quad \text{for all} \;\; \mathbf{v} \in \mathbf{V}.$$

And thus for $t \in [0,T]$ we have

$$\frac{d}{dt}(\mathbf{u}(t), \mathbf{v})_{L^2} + a(\mathbf{u}(t), \mathbf{v}) + b(\mathbf{v}, p(t)) = (\mathbf{g}, \mathbf{v})_{L^2} \quad \text{for all} \;\; \mathbf{v} \in \mathbf{V},$$

i.e., the first equation in (1.47) holds, too. Using the inf-sup property (1.37) we obtain

$$\|p(t)\|_{L^2} \leq c \sup_{\mathbf{v} \in \mathbf{V}} \frac{(p(t), \operatorname{div} \mathbf{v})_{L^2}}{\|\mathbf{v}\|_1} = c \sup_{\mathbf{v} \in \mathbf{V}} \frac{\ell(t)(\mathbf{v})}{\|\mathbf{v}\|_1} = c \, \|\ell(t)\|_{\mathbf{V}'}.$$

Using $\ell \in L^2(0,T,\mathbf{V}')$ it follows that $p \in L^2(0,T;Q)$ holds. Thus $(\mathbf{u}, p)$ is a solution of (1.47). □

Note that in Theorem 1.3.11 the assumptions on the data $\mathbf{g}$ and $\mathbf{u}_0$ are somewhat stronger than in Theorem 1.3.10.

Application to non-stationary Navier-Stokes equations

We now turn to the weak formulation of the non-stationary Navier-Stokes equations (in dimensionless formulation)

$$\begin{aligned} \frac{\partial \mathbf{u}}{\partial t} - \frac{1}{Re} \Delta \mathbf{u} + (\mathbf{u} \cdot \nabla)\mathbf{u} + \nabla p = \mathbf{f} \;\text{ in }\; \Omega \\ \operatorname{div} \mathbf{u} = 0 \;\text{ in }\; \Omega, \end{aligned} \tag{1.48}$$

with a homogeneous Dirichlet boundary condition, $\mathbf{u} = 0$ on $\partial\Omega$, and an initial condition $\mathbf{u}(x,0) = \mathbf{u}_0(x)$ on Ω. The Reynolds number Re is a given strictly positive constant. The weak formulation of this problem takes a form very similar to that of the Stokes equations discussed above. The analysis of well-posedness of this weak formulation, however, is much more complicated as for the Stokes case. Due to the *non*linearity of the Navier-Stokes equations the weak formulation cannot be analyzed in the abstract framework of (1.42) and Theorem 1.3.7. Below we will only present a few main results. For proofs we refer to the literature, e.g. chapter III in [39].

First we introduce a slight generalization of the weak derivative as defined in Definition 1.3.5 (still not using the concept of distributions). For a function $g \in L^2(0,T)$ a weak derivative $h = g' \in L^1(0,T)$ (instead of $\in L^2(0,T)$!) is still well-defined by the condition in (1.38). This is due to the fact that for $\phi \in C_0^\infty(0,T)$, $h \in L^1(0,T)$ we have

$$\left| \int_0^T h(t)\phi(t)\,dt \right| \leq \|\phi\|_{\infty,[0,T]} \int_0^T |h(t)|\,dt = \|\phi\|_{\infty,[0,T]} \|h\|_{L^1} < \infty.$$

Similarly, for $u \in L^2(0,T;V)$ (V a Hilbert space) we have a well-defined weak derivative $u' \in L^1(0,T;V')$ if in Definition 1.3.5 we replace "$w \in L^2(0,T;V')$" by "$w \in L^1(0,T;V')$". For this weak derivative the identity $\frac{d}{dt}(u(t),v)_H = u'(t)(v)$ as in (1.40) still holds. Instead of the space $W^1(0,T;V) = \left\{ v \in L^2(0,T;V) : v' \in L^2(0,T;V') \text{ exists} \right\}$, with V a Hilbert space, we need the larger space

$$W^1_*(0,T;V) := \left\{ v \in L^2(0,T;V) : v' \in L^1(0,T;V') \text{ exists} \right\}$$

which is a Banach space for $\|u\|_{W^1_*(0,T;V)} := \left(\|u\|^2_{L^2(0,T;V)} + \|u'\|^2_{L^1(0,T;V')} \right)^{\frac{1}{2}}$.

For the weak formulation of the Navier-Stokes equations we use the same Gelfand triple (1.45) as for the Stokes problem with spaces $\mathbf{H}_{\mathrm{div}}$, $\mathbf{V}_{\mathrm{div}}$ as in (1.43)-(1.44). We use the same bilinear forms $a(\cdot,\cdot)$, $b(\cdot,\cdot)$ as for the Stokes problem and the trilinear form $c(\mathbf{w};\mathbf{u},\mathbf{v})$ as defined in (1.33c). The following weak formulation of (1.48) is similar to the weak Stokes problem (1.46) :

Determine $\mathbf{u} \in W^1_*(0,T;\mathbf{V}_{\mathrm{div}})$ such that

$$\begin{aligned} \frac{d}{dt}(\mathbf{u}(t),\mathbf{v})_{L^2} + a(\mathbf{u}(t),\mathbf{v}) + c(\mathbf{u}(t);\mathbf{u}(t),\mathbf{v}) &= (\mathbf{g},\mathbf{v})_{L^2} \ \forall\, \mathbf{v} \in \mathbf{V}_{\mathrm{div}}, \\ \mathbf{u}(0) &= \mathbf{u}_0. \end{aligned} \tag{1.49}$$

We comment on some properties of this weak formulation:

- For $\mathbf{g} \in L^2(0,T;\mathbf{V}_{\mathrm{div}})$, $\mathbf{u}_0 \in \mathbf{H}_{\mathrm{div}}$, there *exists* a solution $\mathbf{u}$ of the weak formulation (1.49). This solution has sufficient regularity such that the initial condition $\mathbf{u}(0) = \mathbf{u}_0$ is well-defined.
- *Uniqueness* of this solution is an open problem. Uniqueness of $\mathbf{u}$ can be shown to hold in special cases, for example:
  - If the boundary $\partial\Omega$ is sufficiently smooth, $\mathbf{u}_0 \in \mathbf{V}_{\mathrm{div}}$, $\mathbf{g} \in L^\infty(0,T;\mathbf{H}_{\mathrm{div}})$ then $\mathbf{u}(t)$ is unique on a time interval $[0,T^*]$ with T^* *sufficiently small*.
  - If the Reynolds number Re is sufficiently small, $\mathbf{u}_0$ and $\mathbf{g}$ are sufficiently smooth and these data are sufficiently small (in appropriate norms) then $\mathbf{u}(t)$ is unique for all $t \in [0,T]$.
  - *If* the weak solution $\mathbf{u}$ is sufficiently smooth ($\mathbf{u} \in L^\infty(0,T;\mathbf{H}_{\mathrm{div}}) \cap L^8(0,T;L^4(\Omega))$) then $\mathbf{u}$ is unique. It is not known, however, whether in general this smoothness property holds.

  There is an extensive literature on this topic of uniqueness of a weak solution in the three-dimensional case (i.e. $\Omega \subset \mathbb{R}^3$). Many other special cases are known in the literature. The general case, however, is still unsolved.
- If one considers the weak formulation (1.49) in the *two*-dimensional case (i.e. $\Omega \subset \mathbb{R}^2$) then *existence and uniqueness* have been proved.

Along the same lines as in Lemma 1.3.9 one can show that if the strong formulation (1.48) has a solution $(\mathbf{u},p)$ that is sufficiently smooth ($\mathbf{u} \in C^1\big([0,T];C^2(\overline{\Omega})^3\big)$) then $\mathbf{u}$ is also a solution of (1.49).

In the weak formulation (1.49) the incompressibility condition is fulfilled since it holds for all functions in the space $W^1_*(0,T;\mathbf{V}_{\rm div})$. As for the Stokes problem, a weak formulation in the larger velocity space $W^1_*(0,T;\mathbf{V})$ can be derived in which the incompressibility condition is enforced by using a pressure variable. This weak formulation is as follows:

Determine $\mathbf{u} \in W^1_*(0,T;\mathbf{V}),\ p \in L^2(0,T;Q)$ such that for all $\mathbf{v} \in \mathbf{V},\ q \in Q$:

$$\begin{aligned} \frac{d}{dt}(\mathbf{u}(t),\mathbf{v})_{L^2} + a(\mathbf{u}(t),\mathbf{v}) + c(\mathbf{u}(t);\mathbf{u}(t),\mathbf{v}) + b(\mathbf{v},p(t)) &= (\mathbf{g},\mathbf{v})_{L^2}, \\ b(\mathbf{u}(t),q) &= 0, \\ \mathbf{u}(0) &= \mathbf{u}_0. \end{aligned} \tag{1.50}$$

As in the proof of Theorem 1.3.11 one can show that if $(\mathbf{u},p)$ is a solution of (1.50) then $\mathbf{u}$ is a solution of (1.49). It can be shown that if the solution $\mathbf{u}$ of (1.50) is assumed to be sufficiently smooth then a pressure $p \in L^2(0,T;Q)$ exists such that the pair $(\mathbf{u},p)$ is a solution of (1.50). For a treatment of this topic and a discussion of other similar weak formulations of the Navier-Stokes equations we refer to the literature, e.g. [39, 31].

2

Finite element semi-discretization of the Navier-Stokes equations

2.1 Hood-Taylor finite element spaces

In this section we treat the main topics related to the popular class of so-called Hood-Taylor finite elements for the (spatial) discretization of one phase incompressible flow problems. We introduce these spaces for the d-dimensional case, $d \leq 3$. In our applications we are particularly interested in $d = 3$.

2.1.1 Simplicial finite element spaces

Let Ω be a domain in $\mathbb{R}^d$ and $\mathcal{T}_h = \{T\}$ a subdivision (or triangulation) of Ω in a finite number of simplices T. This triangulation is called *consistent* if the following holds:

1. $\cup_{T\in\mathcal{T}_h} T = \overline{\Omega}$,
2. $\operatorname{int} T_1 \cap \operatorname{int} T_2 = \emptyset$ for all $T_1, T_2 \in \mathcal{T}_h,\ T_1 \neq T_2$,
3. any $(d-1)-$dimensional subsimplex of any $T_1 \in \mathcal{T}_h$ is either a subset of $\partial\Omega$ or a subsimplex of another $T_2 \in \mathcal{T}_h$.

Let $\{\mathcal{T}_h\}$ be a family of triangulations and $h_T := \operatorname{diam}(T)$ for $T \in \mathcal{T}_h$. The index parameter h of $\mathcal{T}_h$ is taken such that

$$h = \max\{\ h_T : T \in \mathcal{T}_h\ \}.$$

Furthermore, for $T \in \mathcal{T}_h$ we define

$$\rho_T := \sup\{\operatorname{diam}(B) :\ B \text{ is a ball contained in } T\ \},$$

A family of consistent triangulations $\{\mathcal{T}_h\}$ is called *regular* if

1. The parameter h approaches zero: $\inf\{\ h : \mathcal{T}_h \in \{\mathcal{T}_h\}\ \} = 0$,
2. $\exists\ \sigma\ :\ \ \frac{h_T}{\rho_T} \leq \sigma$ for all $T \in \mathcal{T}_h$ and all $\mathcal{T}_h \in \{\mathcal{T}_h\}$.

The space of polynomials in $\mathbb{R}^d$ of degree less than or equal $k \geq 0$ is denoted by $\mathcal{P}_k$, i.e., $p \in \mathcal{P}_k$ is of the form

$$p(x) = \sum_{|\alpha|\leq k} \gamma_\alpha x_1^{\alpha_1} x_2^{\alpha_2} \dots x_d^{\alpha_d}, \qquad \gamma_\alpha \in \mathbb{R}.$$

The dimension of $\mathcal{P}_k$ is

$$\dim \mathcal{P}_k = \binom{d+k}{k}. \tag{2.1}$$

The spaces of *simplicial finite elements* are given by

$$\mathbb{X}_h^0 := \{\ v \in L^2(\Omega) :\ v_{|T} \in \mathcal{P}_0 \ \text{ for all } \ T \in \mathcal{T}_h\ \}, \tag{2.2a}$$

$$\mathbb{X}_h^k := \{\ v \in C(\overline{\Omega}) :\ v_{|T} \in \mathcal{P}_k \ \text{ for all } \ T \in \mathcal{T}_h\ \}, \quad k \geq 1. \tag{2.2b}$$

These spaces consist of *piecewise polynomials* which, for $k \geq 1$, are continuous on Ω.

Remark 2.1.1 One can show that $\mathbb{X}_h^k \subset H^1(\Omega)$ holds for all $k \geq 1$.

We will also need simplicial finite element spaces with functions that are zero on $\partial\Omega$:

$$\mathbb{X}_{h,0}^k := \mathbb{X}_h^k \cap H_0^1(\Omega), \quad k \geq 1. \tag{2.3}$$

The values of $v_{|T} \in \mathcal{P}_k$ can be determined by using suitable interpolation points in the simplex T. For this it is convenient to use barycentric coordinates:

Definition 2.1.2 Let T be a non-degenerate d-simplex and $a_j \in \mathbb{R}^d,\ j = 1, \ldots, d+1$ its vertices. Then T can be described by

$$T = \left\{ \sum_{j=1}^{d+1} \lambda_j a_j : 0 \leq \lambda_j \leq 1 \ \ \forall\, j,\ \sum_{j=1}^{d+1} \lambda_j = 1 \right\}. \tag{2.4}$$

To every $x \in T$ there corresponds a unique $(n+1)$-tuple $(\lambda_1, \ldots, \lambda_{n+1})$ as in (2.4). These $\lambda_j,\ 1 \leq j \leq d+1$, are called the *barycentric coordinates* of $x \in T$. The mapping $x \to (\lambda_1, \ldots, \lambda_{d+1})$ is affine.

Using these barycentric coordinates we define the set

$$L_k(T) := \left\{ \sum_{j=1}^{d+1} \lambda_j a_j : \lambda_j \in \{0, \frac{1}{k}, \ldots, \frac{k-1}{k}, 1\} \ \ \forall\, j,\ \sum_{j=1}^{d+1} \lambda_j = 1 \right\}$$

which is called the principal lattice of order k (in T). Examples for $d = 3$ and $k = 1, 2$ are given in Figure 2.1.

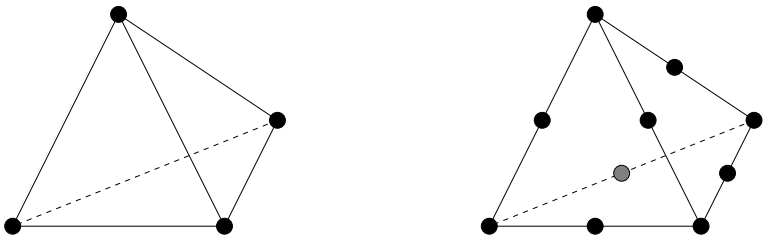

Fig. 2.1. Principal lattice, $d = 3$, $k = 1$ (left) and $k = 2$ (right).

This principal lattice consists of $\begin{pmatrix} d+k \\ k \end{pmatrix}$ points (cf. (2.1)) and these can be used to determine a unique polynomial $p \in \mathcal{P}_k$:

Lemma 2.1.3 *Let T be a non-degenerated d-simplex. Then any polynomial $p \in \mathcal{P}_k$ is uniquely determined by its values on the principal lattice $L_k(T)$.*

Using this lattice, for $u \in C(\overline{\Omega})$ we define a corresponding function $I_{\mathbb{X}}^k u \in L^2(\Omega)$ by piecewise polynomial interpolation on each simplex $T \in \mathcal{T}_h$:

$$\forall\, T \in \mathcal{T}_h : \ (I_{\mathbb{X}}^k u)_{|T} \in \mathcal{P}_k \ \text{ such that } \ (I_{\mathbb{X}}^k u)(x_j) = u(x_j) \ \ \forall\, x_j \in L_k(T). \tag{2.5}$$

The piecewise polynomial function $I_{\mathbb{X}}^k u$ is continuous on Ω:

Lemma 2.1.4 *For $k \geq 1$ and $u \in C(\overline{\Omega})$ we have $I_{\mathbb{X}}^k u \in \mathbb{X}_h^k$.*

Proof. We consider the case $d = 3$, $k = 2$, cf. Figure 2.1 (right). By definition we have $I^2_{\mathbb{X}} u \in \mathcal{P}_2$, thus we only have to show that $I^2_{\mathbb{X}} u$ is continuous across triangular faces between adjacent tetrahedra T_1, T_2. Define $p_i := (I^2_{\mathbb{X}} u)_{|T_i}$, $i = 1, 2$. At the six interpolation points x_j, $j = 1, \ldots, 6$, on the face $T_1 \cap T_2$ we have $p_1(x_j) = p_2(x_j) = u(x_j)$. The functions $(p_i)_{|T_1 \cap T_2}$ are two-dimensional polynomials of degree 2, which are uniquely determined by the six values $p_i(x_j)$. We conclude that $p_1 = p_2$ on $T_1 \cap T_2$ and thus $I^2_{\mathbb{X}} u$ is continuous across $T_1 \cap T_2$. Similar arguments can be applied to prove the result for the general case. □

Using the embedding result $H^m(\Omega) \hookrightarrow C(\overline{\Omega})$ for $m > \frac{d}{2}$, cf. (1.28), we obtain the following corollary.

Corollary 2.1.5 For $k \geq 1, m \geq 2$ we have:

$$\begin{aligned} I^k_{\mathbb{X}} u \in \mathbb{X}^k_h \quad &\text{for all} \quad u \in H^m(\Omega), \\ I^k_{\mathbb{X}} u \in \mathbb{X}^k_{h,0} \quad &\text{for all} \quad u \in H^m(\Omega) \cap H^1_0(\Omega). \end{aligned}$$

Using interpolation error bounds one can derive the following main result.

Theorem 2.1.6 *Let $\{\mathcal{T}_h\}$ be a regular family of triangulations of Ω consisting of d-simplices and let $\mathbb{X}^k_h$ be the corresponding finite element space as in* (2.2b)*. For $2 \leq m \leq k+1$ and $t \in \{0, 1\}$ the following holds:*

$$\|u - I^k_{\mathbb{X}} u\|_t \leq C h^{m-t} |u|_m \quad \textit{for all} \quad u \in H^m(\Omega). \tag{2.6}$$

Recall that $\|\cdot\|_t$ and $|\cdot|_m$ are the norm on $H^t(\Omega)$ and the semi-norm on $H^m(\Omega)$, respectively, and that $\|\cdot\|_0 = \|\cdot\|_{L^2}$. A proof of this result can be found in many textbooks on finite element methods, e.g. [11, 5, 6, 16]. As a direct consequence of this interpolation error bound we obtain the following approximation error bound.

Theorem 2.1.7 *Let $\{\mathcal{T}_h\}$ be a regular family of triangulations of Ω consisting of d-simplices and let $\mathbb{X}^k_h$, $\mathbb{X}^k_{h,0}$ be the corresponding finite element space as in* (2.2b)*,* (2.3)*. For $2 \leq m \leq k+1$ and $t \in \{0, 1\}$ the following holds:*

$$\inf_{v_h \in \mathbb{X}^k_h} \|u - v_h\|_t \leq C h^{m-t} |u|_m \ \textit{for all} \ \ u \in H^m(\Omega), \tag{2.7a}$$

$$\inf_{v_h \in \mathbb{X}^k_{h,0}} \|u - v_h\|_t \leq C h^{m-t} |u|_m \ \textit{for all} \ \ u \in H^m(\Omega) \cap H^1_0(\Omega). \tag{2.7b}$$

A function $u \in H^1(\Omega)$ is not necessarily continuous and therefore it may be that the nodal interpolation $I^k_{\mathbb{X}} u$ is not well-defined. Other quasi-interpolation operators (e.g. so-called Clement interpolation) have been developed that are well-defined for $u \in H^1(\Omega)$. Using these one can prove the approximation result

$$\inf_{v_h \in \mathbb{X}^k_h} \|u - v_h\|_{L^2} \leq C h |u|_1 \text{ for all } \ u \in H^1(\Omega), \quad k \geq 0. \tag{2.8}$$

2.1.2 Hood-Taylor finite element discretization of the Oseen problem

In this section we use the simplicial finite element spaces $\mathbb{X}^k_h$ introduced above for the discretization of the variational Oseen problem in (1.34). We apply the abstract results on the Galerkin discretization of saddle point problems given in the Appendix, Section 4.3, to the variational Oseen problem (1.34). Let $\mathbf{V}_h$ and Q_h be finite dimensional subspaces of $\mathbf{V} = H^1_0(\Omega)^3$ and $Q = L^2_0(\Omega)$, respectively. Then the *Galerkin discretization of the Oseen problem* (1.34) is as follows:

Find $\mathbf{u}_h \in \mathbf{V}_h$ and $p_h \in Q_h$, such that

$$\begin{aligned} \xi m(\mathbf{u}_h, \mathbf{v}_h) + a(\mathbf{u}_h, \mathbf{v}_h) + c(\mathbf{u}_h, \mathbf{v}_h) + b(\mathbf{v}_h, p_h) &= (\mathbf{g}, \mathbf{v}_h)_{L^2} \qquad \forall\ \mathbf{v}_h \in \mathbf{V}_h \\ b(\mathbf{u}_h, q_h) &= 0 \qquad \forall\ q_h \in Q_h. \end{aligned} \tag{2.9}$$

This is of the form (4.20) with bilinear forms

$$\begin{aligned} \hat{a}(\mathbf{u}, \mathbf{v}) &= \xi m(\mathbf{u}, \mathbf{v}) + a(\mathbf{u}, \mathbf{v}) + c(\mathbf{u}, \mathbf{v}) \quad \text{on} \quad \mathbf{V} \times \mathbf{V}, \\ \hat{b}(\mathbf{u}, q) &= b(\mathbf{u}, q) \quad \text{on} \quad \mathbf{V} \times Q. \end{aligned} \tag{2.10}$$

The right-hand sides are given by $f_1(\mathbf{v}) = (\mathbf{g}, \mathbf{v})_{L^2}$, $f_2 = 0$. We recall the definitions of the bilinear forms:

$$m(\mathbf{u}, \mathbf{v}) = \int_\Omega \mathbf{u} \cdot \mathbf{v}\, dx, \quad a(\mathbf{u}, \mathbf{v}) = \frac{1}{Re} \int_\Omega \nabla \mathbf{u} \cdot \nabla \mathbf{v}\, dx$$
$$c(\mathbf{u}, \mathbf{v}) = c(\mathbf{w}; \mathbf{u}, \mathbf{v}) = \int_\Omega (\mathbf{w} \cdot \nabla \mathbf{u}) \cdot \mathbf{v}\, dx, \quad b(\mathbf{u}, q) = -\int_\Omega q \operatorname{div} \mathbf{u}\, dx.$$

The bilinear form $\hat{b}(\cdot, \cdot)$ satisfies the inf-sup condition (4.21a) in Theorem 4.3.5, cf. (1.37) in Section 1.3.2. We make the assumption $\xi - \frac{1}{2} \operatorname{div} \mathbf{w} \geq 0$ on Ω, also used in Theorem 1.3.4. Then the bilinear form $\hat{a}(\cdot, \cdot)$ satisfies the ellipticity condition (4.21b) in Theorem 4.3.5. In view of the *discrete inf-sup condition* (4.21c) we introduce the following definition.

Definition 2.1.8 The pair $(\mathbf{V}_h, Q_h)$ is called *stable* if there exists a constant $\hat{\beta} > 0$ *independent* of h such that

$$\sup_{\mathbf{v}_h \in \mathbf{V}_h} \frac{\hat{b}(\mathbf{v}_h, q_h)}{\|\mathbf{v}_h\|_1} \geq \hat{\beta}\, \|q_h\|_{L^2} \quad \text{for all} \quad q_h \in Q_h. \tag{2.11}$$

□

In the literature this is also often called the *LBB stability* condition of the finite element pair $(\mathbf{V}_h, Q_h)$ (due to Ladyzenskaya, Babuska, Brezzi).

LBB-stable pairs of finite element spaces

We now turn to the question which pairs of finite element spaces can be used for the Galerkin discretization of the Oseen problem. In view of the simplicial finite element spaces introduced in Section 2.1.1 we consider the following so-called *Hood-Taylor* pair:

$$\big((\mathbb{X}_{h,0}^k)^d\,,\, \mathbb{X}_h^{k-1} \cap L_0^2(\Omega)\big), \quad k \geq 1. \tag{2.12}$$

The following remark shows that the issue of LBB-stability needs a careful analysis.

Remark 2.1.9 Take $d = 2$, $\Omega = (0,1)^2$ and a uniform triangulation $\mathcal{T}_h$ of Ω that is defined as follows. For $N \in \mathbb{N}$ and $h := \frac{1}{N+1}$ the domain Ω is subdivided in squares with sides of length h and vertices in the set $\{\, (ih, jh) : 0 \leq i, j \leq N+1 \,\}$. The triangulation $\mathcal{T}_h$ is obtained by subdividing every square in two triangles by inserting a diagonal from (ih, jh) to $((i+1)h, (j+1)h)$. For the pair $(\mathbf{V}_h, Q_h)$ we take

$$(\mathbf{V}_h, Q_h) := \big((\mathbb{X}_{h,0}^1)^2\,,\, \mathbb{X}_h^0 \cap L_0^2(\Omega)\big).$$

The space $\mathbf{V}_h$ has dimension $2N^2$ and $\dim(Q_h) = 2(N+1)^2 - 1$. From $\dim(\mathbf{V}_h) < \dim(Q_h)$ and Remark 4.3.6 it follows that the condition (2.11) does not hold.
The same argument applies to the three dimensional case with a uniform triangulation of $(0,1)^3$ consisting of tetrahedra (every cube is subdivided in 6 tetrahedra). In this case we have $\dim(\mathbf{V}_h) = 3N^3$ and $\dim(Q_h) = 6(N+1)^3 - 1$.

This remark implies that in general for $k = 1$ the Hood-Taylor pair in (2.12) is *not* LBB stable. However, for $k \geq 2$ this pair is stable:

Theorem 2.1.10 *Let $\{\mathcal{T}_h\}$ be a regular family of triangulations consisting of simplices. We assume that every $T \in \mathcal{T}_h$ has at least one vertex in the interior of Ω. Then the Hood-Taylor pair of finite element spaces with $k \geq 2$ is LBB stable.*

For a proof of this important result we refer to the literature, [3, 4, 7]. Using this stability property of the Hood-Taylor finite element spaces the following discretization error bound can be derived.

Theorem 2.1.11 *Let $\{\mathcal{T}_h\}$ be a regular family of triangulations as in Theorem* 2.1.10. *Consider the discrete Oseen problem* (2.9) *with Hood-Taylor finite element spaces as in* (2.12), *$k \geq 2$. Suppose that $\xi - \frac{1}{2}\operatorname{div}\mathbf{w} \geq 0$ holds and that the continuous solution $(\mathbf{u}, p)$ lies in $H^m(\Omega)^3 \times H^{m-1}(\Omega)$ with $m \geq 2$. For $2 \leq m \leq k+1$ the following holds:*

$$\|\mathbf{u} - \mathbf{u}_h\|_1 + \|p - p_h\|_{L^2} \leq C\, h^{m-1}\big(|\mathbf{u}|_m + |p|_{m-1}\big)$$

with a constant C independent of h and of $(\mathbf{u}, p)$.

Proof. We consider the bilinear forms $\hat{a}(\cdot,\cdot)$ and $\hat{b}(\cdot,\cdot)$ as in (2.10) and apply Theorem 4.3.5. Due to the inf-sup property (1.37) and the assumption $\xi - \frac{1}{2}\operatorname{div}\mathbf{w} \geq 0$ the conditions (4.21a) and (4.21b) are satisfied (cf. Theorem 1.3.4). Due to the LBB stability of the Hood-Taylor pair (2.12) with $k \geq 2$ the discrete inf-sup condition (4.21c) is fulfilled with a constant β_h *in*dependent of h. For the approximation error corresponding to the Hood-Taylor finite element spaces the results in Theorem 2.1.7 and in (2.8) can be used:

$$\inf_{v_h \in \mathbb{X}_{h,0}^k} \|\mathbf{u} - \mathbf{v}_h\|_1 \leq Ch^{m-1}|\mathbf{u}|_m,$$
$$\inf_{q_h \in \mathbb{X}_h^{k-1}} \|p - q_h\|_{L^2} \leq Ch^{m-1}|p|_{m-1}.$$

This completes the proof. □

Note that in this theorem sufficient *regularity* of the Oseen problem is required, namely that the solution $(\mathbf{u}, p)$ lies in the space $H^m(\Omega)^3 \times H^{m-1}(\Omega)$ with $m \geq 2$. For a discussion of this regularity issue we refer to the literature. If this regularity assumption holds both for the Oseen problem (1.34) and for the corresponding adjoint problem, i.e. a problem as in (1.34) with $\mathbf{w}$ replaced by $-\mathbf{w}$, then a discretization error bound

$$\|\mathbf{u} - \mathbf{u}_h\|_{L^2} \leq C\, h^m\big(|\mathbf{u}|_m + |p|_{m-1}\big) \tag{2.13}$$

can be shown to hold, with C independent of h and of $(\mathbf{u}, p)$.

Remark 2.1.12 For the (generalized) Stokes equations we have $\mathbf{w} = 0$ and thus the problem is symmetric, i.e. the adjoint problem equals the original one. It is known ([30, 10, 13]) that for this case the regularity property $(\mathbf{u}, p) \in H^m(\Omega)^3 \times H^{m-1}(\Omega)$, with $m \geq 2$, is satisfied for $m = 2$ if Ω is convex and for the general case $m \geq 2$ if $\partial\Omega$ is sufficiently smooth.

2.1.3 Matrix-vector representation of the discrete problem

Consider the variational discrete Oseen problem (2.9) with Hood-Taylor finite element spaces, $(\mathbf{V}_h, Q_h) = \big((\mathbb{X}_{h,0}^k)^3\,,\, \mathbb{X}_h^{k-1} \cap L_0^2(\Omega)\big)$, $k \geq 2$. For computing the unique discrete solution $(\mathbf{u}_h, p_h)$ we introduce the *nodal* basis functions in the simplicial finite element space $\mathbb{X}_h^k$, which are defined as follows. The union of all lattice points in $L_k(T)$, $T \in \mathcal{T}_h$, form the set of grid points $(\mathbf{x}_i)_{1 \leq i \leq K}$. Note that $\dim(\mathbb{X}_h^k) = K$. To each of these grid points there corresponds

a nodal finite element function $\phi_i \in \mathbb{X}_h^k$ with the property $\phi_i(\mathbf{x}_i) = 1,\ \phi_i(\mathbf{x}_j) = 0$ for all $j \neq i$. The set of functions $(\phi_i)_{1\leq i\leq K}$ forms the nodal basis of $\mathbb{X}_h^k$. In case of $\mathbb{X}_{h,0}^k$ only the nodal functions corresponding to grid points in the interior of Ω are used. In case of vector functions. i.e. $(\mathbb{X}_{h,0}^k)^d$ with $d > 1$, to each grid point there correspond d of such nodal functions, namely one for each of the d components in the vector function. Let $\{\boldsymbol{\xi}_i\}_{1\leq i\leq N}$ and $\{\psi_i\}_{1\leq i\leq K}$ be such nodal bases of the finite element spaces $\mathbf{V}_h = (\mathbb{X}_{h,0}^k)^3$ and $\mathbb{X}_h^{k-1}$, respectively. Hence, $\dim(\mathbb{X}_{h,0}^k)^3 = N,\ \dim(\mathbb{X}_h^{k-1}) = K$. Consider the representations

$$\mathbf{u}_h = \sum_{j=1}^{N} u_j \boldsymbol{\xi}_j, \quad \vec{\mathbf{u}} := (u_1, \ldots, u_N) \tag{2.14}$$

$$p_h = \sum_{j=1}^{K} p_j \psi_j, \quad \vec{\mathbf{p}} := (p_1, \ldots, p_K). \tag{2.15}$$

Using this the discrete Oseen problem (2.9) can be reformulated as follows:

Determine $\vec{\mathbf{u}} \in \mathbb{R}^N,\ \vec{\mathbf{p}} \in \mathbb{R}^K$ with $(p_h, 1)_{L^2} = 0$ such that

$$\begin{pmatrix} \xi\mathbf{M} + \mathbf{A} + \mathbf{C} & \mathbf{B}^T \\ \mathbf{B} & 0 \end{pmatrix} \begin{pmatrix} \vec{\mathbf{u}} \\ \vec{\mathbf{p}} \end{pmatrix} = \begin{pmatrix} \vec{\mathbf{f}} \\ 0 \end{pmatrix} \tag{2.16}$$

where

$$\mathbf{M} \in \mathbb{R}^{N\times N}, \quad \mathbf{M}_{ij} = \int_\Omega \boldsymbol{\xi}_i \cdot \boldsymbol{\xi}_j \, dx \tag{2.17a}$$

$$\mathbf{A} \in \mathbb{R}^{N\times N}, \quad \mathbf{A}_{ij} = \frac{1}{Re} \int_\Omega \nabla\boldsymbol{\xi}_i \cdot \nabla\boldsymbol{\xi}_j \, dx \tag{2.17b}$$

$$\mathbf{C} = \mathbf{C}(\mathbf{w}) \in \mathbb{R}^{N\times N}, \quad \mathbf{C}_{ij} = \int_\Omega (\mathbf{w} \cdot \nabla\boldsymbol{\xi}_j) \cdot \boldsymbol{\xi}_i \, dx \tag{2.17c}$$

$$\mathbf{B} \in \mathbb{R}^{K\times N}, \quad \mathbf{B}_{ij} = -\int_\Omega \psi_i \operatorname{div} \boldsymbol{\xi}_j \, dx \tag{2.17d}$$

$$\vec{\mathbf{f}} \in \mathbb{R}^N, \quad \vec{\mathbf{f}}_i = \int_\Omega \mathbf{g} \cdot \boldsymbol{\xi}_i \, dx. \tag{2.17e}$$

$\mathbf{M}, \mathbf{A}, \mathbf{C}$ are called mass matrix, diffusion matrix and convection matrix, respectively. Matrices with a block structure as in (2.16) are called saddle point matrices.

2.1.4 Hood-Taylor semi-discretization of the non-stationary (Navier-)Stokes problem

We recall the weak formulation of the non-stationary Stokes equations given (1.47): Find $\mathbf{u} \in W^1(0,T;\mathbf{V})$ and $p \in L^2(0,T;Q)$, such that $\mathbf{u}(0) = \mathbf{u}_0$ and

$$\begin{aligned} \frac{d}{dt} m(\mathbf{u}(t), \mathbf{v}) + a(\mathbf{u}(t), \mathbf{v}) + b(\mathbf{v}, p(t)) &= (\mathbf{g}, \mathbf{v})_{L^2} \quad \forall\ \mathbf{v} \in \mathbf{V} \\ b(\mathbf{u}(t), q) &= 0 \quad \forall\ q \in Q, \end{aligned} \tag{2.18}$$

for almost all $t \in [0,T]$. For the *spatial* discretization of this problem we use the Galerkin approach with a stable Hood-Taylor pair from (2.12):

$$(\mathbf{V}_h, Q_h) = \left((\mathbb{X}_{h,0}^k)^3, \mathbb{X}_h^{k-1} \cap L_0^2(\Omega)\right), \quad k \geq 2.$$

Let $\mathbf{u}_{0,h} \in \mathbf{V}_h$ be an approximation of the initial condition $\mathbf{u}_0$. The Galerkin *semi*-discretization reads: Find $\mathbf{u}_h(t) \in \mathbf{V}_h$, with $\mathbf{u}_h(0) = \mathbf{u}_{0,h}$, and $p_h(t) \in Q_h$ such that:

$$\begin{aligned} \frac{d}{dt} m(\mathbf{u}_h(t), \mathbf{v}_h) + a(\mathbf{u}_h(t), \mathbf{v}_h) + b(\mathbf{v}_h, p_h(t)) &= (\mathbf{g}, \mathbf{v}_h)_{L^2} \ \forall\ \mathbf{v}_h \in \mathbf{V}_h, \\ b(\mathbf{u}_h(t), q_h) &= 0 \ \ \forall\ q_h \in Q_h, \end{aligned} \tag{2.19}$$

for all $t \in [0,T]$. Let $\{\boldsymbol{\xi}_i\}_{1\le i\le N}$ and $\{\psi_i\}_{1\le i\le K}$ be the standard nodal bases of the finite element spaces $\mathbf{V}_h$, $\mathbb{X}_h^{k-1}$ and consider the representations

$$\mathbf{u}_h(t) = \sum_{j=1}^{N} u_j(t)\boldsymbol{\xi}_j, \quad \vec{\mathbf{u}}(t) := (u_1(t), \dots, u_N(t)) \tag{2.20a}$$

$$\mathbf{u}_{0,h} = \sum_{j=1}^{N} u_{0,j}\boldsymbol{\xi}_j, \quad \vec{\mathbf{u}}_0 := (u_{0,1}, \dots, u_{0,N}) \tag{2.20b}$$

$$p_h(t) = \sum_{j=1}^{K} p_j(t)\psi_j, \quad \vec{\mathbf{p}}(t) := (p_1(t), \dots, p_K(t)). \tag{2.20c}$$

Using this the Galerkin discretization can be rewritten as

Determine $\vec{\mathbf{u}}(t) \in \mathbb{R}^N,\ \vec{\mathbf{p}}(t) \in \mathbb{R}^K$ with $\vec{\mathbf{u}}(0) = \vec{\mathbf{u}}_0$ and $(p_h(t), 1)_{L^2} = 0$ such that

$$\begin{aligned} \mathbf{M}\frac{d\vec{\mathbf{u}}}{dt}(t) + \mathbf{A}\vec{\mathbf{u}}(t) + \mathbf{B}^T\vec{\mathbf{p}}(t) &= \vec{\mathbf{f}} \\ \mathbf{B}\vec{\mathbf{u}}(t) &= 0, \end{aligned} \tag{2.21}$$

for all $t \in [0,T]$,

with $\mathbf{M}, \mathbf{A}, \mathbf{B}$ and $\vec{\mathbf{f}}$ as in (2.17).

Thus we obtain a system of *differential algebraic equations* (DAEs) for the unknown vector functions $\vec{\mathbf{u}}(t)$, $\vec{\mathbf{p}}(t)$. Time discretization methods for this system are discussed in Chapter 3.

Lemma 2.1.13 *The problem in* (2.19)*, or equivalently* (2.21)*, has a unique solution.*

Proof. A proof is given in Section 3.2, Remark 3.2.1. □

We derive a bound for the discretization error of the semi-discrete Stokes problem in (2.19).

Theorem 2.1.14 *Let* $(\mathbf{u}, p)$ *and* $(\mathbf{u}_h, p_h)$ *be the solution of* (2.18) *and* (2.19)*, respectively. Assume that* $(\mathbf{u}, p)$ *is sufficiently smooth:* $\mathbf{u} \in C^1([0,T]; H^m(\Omega)^3)$, $p \in C^1([0,T]; H^{m-1}(\Omega))$ *and that* $2 \le m \le k+1$*, with* $k \ge 2$*. Furthermore, we assume that for the discretization of the* stationary *Stokes problem with Hood-Taylor finite elements the error bound* (2.13) *holds (cf. Remark 2.1.12). Then the following holds for* $t \in [0,T]$*:*

$$\|\mathbf{u}(t) - \mathbf{u}_h(t)\|_{L^2} \le e^{-c_0 t}\|\mathbf{u}_0 - \mathbf{u}_{0,h}\|_{L^2} + ch^m E_t(\mathbf{u}, p), \tag{2.22}$$

$$|\mathbf{u}(t) - \mathbf{u}_h(t)|_1 \le |\mathbf{u}_0 - \mathbf{u}_{0,h}|_1 + c\big(1 + h\sqrt{t}\big)h^{m-1}E_t(\mathbf{u}, p), \tag{2.23}$$

$$\Big(\int_0^t \|p(\tau) - p_h(\tau)\|_{L^2}^2\, d\tau\Big)^{\frac{1}{2}} \le c|\mathbf{u}_0 - \mathbf{u}_{0,h}|_1 + c\sqrt{t}\, h^{m-1}E_t(\mathbf{u}, p), \tag{2.24}$$

with constants $c_0 > 0$*,* c *independent of* h *and* t *and*

$$E_t(\mathbf{u}, p) := \sum_{\ell=0}^{1} \max_{0\le\tau\le t} \Big(\big|\mathbf{u}^{(\ell)}(\tau)\big|_m + \big|p^{(\ell)}(\tau)\big|_{m-1}\Big).$$

Proof. Take $\mathbf{w} \in \mathbf{V}_{\mathrm{div}} := \{\, \mathbf{v} \in \mathbf{V} : \operatorname{div} \mathbf{v} = 0 \,\}$, $r \in Q$ and define $\ell(\mathbf{v}) := a(\mathbf{w}, \mathbf{v}) + b(\mathbf{v}, r)$, $\mathbf{v} \in \mathbf{V}$. Then $\ell \in \mathbf{V}'$ and $(\mathbf{w}, r)$ is the unique solution of the stationary Stokes problem

$$\begin{aligned} a(\mathbf{w}, \mathbf{v}) + b(\mathbf{v}, r) &= \ell(\mathbf{v}) \quad \text{for all } \ \mathbf{v} \in \mathbf{V}, \\ b(\mathbf{w}, q) &= 0 \quad \text{for all } \ q \in Q. \end{aligned}$$

Let $(\mathbf{w}_h, r_h) \in \mathbf{V}_h \times Q_h$ be the unique solution of the Galerkin discretization of this problem. The mapping $S : (\mathbf{w}, r) \to (\mathbf{w}_h, r_h)$ is linear on $\mathbf{V}_{\mathrm{div}} \times Q$ and for $\mathbf{w}$ and r sufficiently smooth we have

$$\|\mathbf{w} - \mathbf{w}_h\|_1 + \|r - r_h\|_{L^2} \le c h^{m-1} \big(|\mathbf{w}|_m + |r|_{m-1} \big), \tag{2.25}$$

$$\|\mathbf{w} - \mathbf{w}_h\|_{L^2} \le c h^{m} \big(|\mathbf{w}|_m + |r|_{m-1} \big), \tag{2.26}$$

with constants c independent of h and of $(\mathbf{w}, r)$. Let $(\mathbf{u}, p)$ and $(\mathbf{u}_h, p_h)$ be as defined in the theorem. Define, for $t \in [0, T]$, $(\mathbf{w}_h(t), r_h(t)) := S(\mathbf{u}(t), p(t))$ and

$$\begin{aligned} \mathbf{e}_h(t) := \mathbf{u}(t) - \mathbf{u}_h(t) &= \big(\mathbf{u}(t) - \mathbf{w}_h(t) \big) + \big(\mathbf{w}_h(t) - \mathbf{u}_h(t) \big) =: \rho_h(t) + \theta_h(t), \\ p(t) - p_h(t) &= \big(p(t) - r_h(t) \big) + \big(r_h(t) - p_h(t) \big) =: \xi_h(t) + \eta_h(t). \end{aligned}$$

Note that $\theta_h(t) \in \mathbf{V}_h$, $\eta_h(t) \in Q_h$ and $(\mathbf{w}_h'(t), r_h'(t)) = \frac{d}{dt} S(\mathbf{u}(t), p(t)) = S(\mathbf{u}'(t), p'(t))$. Due to (2.25), (2.26) we have

$$\|\rho_h(t)\|_1 \le c h^{m-1} E_t(\mathbf{u}, p), \tag{2.27}$$

$$\|\rho_h(t)\|_{L^2} \le c h^{m} E_t(\mathbf{u}, p), \tag{2.28}$$

$$\|\rho_h'(t)\|_{L^2} \le c h^{m} E_t(\mathbf{u}, p), \tag{2.29}$$

$$\|\xi_h(t)\|_{L^2} \le c h^{m-1} E_t(\mathbf{u}, p). \tag{2.30}$$

Subtraction of the variational equations for $(\mathbf{u}, p)$ and $(\mathbf{u}_h, p_h)$ results in

$$(\mathbf{e}_h'(t), \mathbf{v}_h)_{L^2} + a(\mathbf{e}_h(t), \mathbf{v}_h) + b(\mathbf{v}_h, p(t) - p_h(t)) = 0 \quad \text{for all } \ \mathbf{v}_h \in \mathbf{V}_h.$$

Using the definitions we obtain

$$a(\rho_h(t), \mathbf{v}_h) + b(\mathbf{v}_h, p(t) - p_h(t)) = b(\mathbf{v}_h, \eta_h(t)) \quad \text{for all } \ \mathbf{v}_h \in \mathbf{V}_h.$$

Using this and the splitting $\mathbf{e}_h(t) = \rho_h(t) + \theta_h(t)$ we get

$$(\theta_h'(t), \mathbf{v}_h)_{L^2} + a(\theta_h(t), \mathbf{v}_h) + b(\mathbf{v}_h, \eta_h(t)) = -(\rho_h'(t), \mathbf{v}_h)_{L^2} \tag{2.31}$$

for all $\mathbf{v}_h \in \mathbf{V}_h$. Based on this fundamental relation the following bounds can be derived:

$$\|\theta_h(t)\|_{L^2} \le e^{-c_0 t} \|\theta_h(0)\|_{L^2} + c h^m E_t(\mathbf{u}, p), \tag{2.32}$$

$$|\theta_h(t)|_1 \le |\theta_h(0)|_1 + c\sqrt{t}\, h^m E_t(\mathbf{u}, p), \tag{2.33}$$

$$\Big(\int_0^t |\theta_h(\tau)|_1^2 \, d\tau \Big)^{\frac{1}{2}} \le c \|\theta_h(0)\|_{L^2} + c\sqrt{t}\, h^m E_t(\mathbf{u}, p), \tag{2.34}$$

$$\Big(\int_0^t \|\theta_h'(\tau)\|_{L^2}^2 \, d\tau \Big)^{\frac{1}{2}} \le c |\theta_h(0)|_1 + c\sqrt{t}\, h^m E_t(\mathbf{u}, p), \tag{2.35}$$

with constants $c_0 > 0$ and c independent of h and t. We now prove these inequalities. In (2.31) we take $\mathbf{v}_h = \theta_h(t)$ and use that $b(\theta_h(t), \eta_h(t)) = 0$, resulting in

$$\frac{1}{2} \frac{d}{dt} \|\theta_h(t)\|_{L^2}^2 + \frac{1}{Re} |\theta_h(t)|_1^2 \le \|\rho_h'(t)\|_{L^2} \|\theta_h(t)\|_{L^2}. \tag{2.36}$$

Using the Poincaré-Friedrichs inequality $\|\theta_h(t)\|_{L^2} \le c |\theta_h(t)|_1$ we obtain from this

$$\frac{1}{2}\frac{d}{dt}\|\theta_h(t)\|_{L^2}^2 + \frac{1}{2}\frac{1}{Re}|\theta_h(t)|_1^2 \le c\|\rho_h'(t)\|_{L^2}^2.$$

Integration over $[0,t]$ results in

$$\begin{aligned}\int_0^t |\theta_h(\tau)|_1^2\,d\tau &\le c\|\theta_h(0)\|_{L^2}^2 + c\int_0^t \|\rho_h'(\tau)\|_{L^2}^2\,d\tau \\ &\le c\|\theta_h(0)\|_{L^2}^2 + c\,t h^{2m}E_t(\mathbf{u},p)^2,\end{aligned}$$

which proves (2.34). From (2.36) we also obtain

$$\|\theta_h(t)\|_{L^2}\frac{d}{dt}\|\theta_h(t)\|_{L^2} + c_0\|\theta_h(t)\|_{L^2}^2 \le \|\rho_h'(t)\|_{L^2}\|\theta_h(t)\|_{L^2},$$

with $c_0 > 0$, and hence

$$\frac{d}{dt}\|\theta_h(t)\|_{L^2} + c_0\|\theta_h(t)\|_{L^2} \le \|\rho_h'(t)\|_{L^2}.$$

Multiplication by $e^{c_0 t}$ and integration over $[0,t]$ yields

$$\begin{aligned}\|\theta_h(t)\|_{L^2} &\le e^{-c_0 t}\|\theta_h(0)\|_{L^2} + \int_0^t e^{-c_0(t-\tau)}\|\rho_h'(\tau)\|_{L^2}\,d\tau \\ &\le e^{-c_0 t}\|\theta_h(0)\|_{L^2} + ch^m E_t(\mathbf{u},p)\int_0^t e^{-c_0(t-\tau)}\,d\tau \\ &\le e^{-c_0 t}\|\theta_h(0)\|_{L^2} + ch^m E_t(\mathbf{u},p),\end{aligned}$$

with $c_0 > 0$ and c independent of h and t. Thus (2.32) holds. In (2.31) we now substitute $\mathbf{v}_h = \theta_h'(t)$. Using $b(\theta_h'(t),\eta_h(t)) = 0$ we get

$$\begin{aligned}\|\theta_h'(t)\|_{L^2}^2 + \frac{1}{2}\frac{1}{Re}\frac{d}{dt}|\theta_h(t)|_1^2 &\le \|\rho_h'(t)\|_{L^2}\|\theta_h'(t)\|_{L^2} \\ &\le \frac{1}{2}\|\rho_h'(t)\|_{L^2}^2 + \frac{1}{2}\|\theta_h'(t)\|_{L^2}^2,\end{aligned}$$

and thus

$$\|\theta_h'(t)\|_{L^2}^2 + \frac{1}{Re}\frac{d}{dt}|\theta_h(t)|_1^2 \le \|\rho_h'(t)\|_{L^2}^2.$$

Integrating this results in

$$\int_0^t \|\theta_h'(\tau)\|_{L^2}^2\,d\tau + \frac{1}{Re}|\theta_h(t)|_1^2 \le \frac{1}{Re}|\theta_h(0)|_1^2 + c\,t h^{2m}E_t(\mathbf{u},p)^2,$$

which proves the results in (2.33) and (2.35).

Using (2.32)-(2.35) we derive the bounds stated in the theorem. Note that $\|\theta_h(0)\|_{L^2} \le \|\mathbf{e}_h(0)\|_{L^2} + \|\rho_h(0)\|_{L^2} \le \|\mathbf{e}_h(0)\|_{L^2} + ch^m E_t(\mathbf{u},p)$ holds, and thus, using (2.28) and (2.32) we get

$$\|\mathbf{e}_h(t)\|_{L^2} \le \|\theta_h(t)\|_{L^2} + \|\rho_h(t)\|_{L^2} \le e^{-c_0 t}\|\mathbf{e}_h(0)\|_{L^2} + ch^m E_t(\mathbf{u},p),$$

hence, the result in (2.22) holds. Similarly, from $|\theta_h(0)|_1 \le |\mathbf{e}_h(0)|_1 + ch^{m-1}E_t(\mathbf{u},p)$ and (2.27), (2.33) we obtain

$$\begin{aligned}|\mathbf{e}_h(t)|_1 &\le |\theta_h(t)|_1 + |\rho_h(t)|_1 \le |\mathbf{e}_h(0)|_1 + c\sqrt{t}h^m E_t(\mathbf{u},p) + ch^{m-1}E_t(\mathbf{u},p) \\ &= |\mathbf{e}_h(0)|_1 + ch^{m-1}\big(1+\sqrt{t}h\big)E_t(\mathbf{u},p),\end{aligned}$$

which proves the result in (2.23). For the pressure error bound we use the LBB stability property of the Hood-Taylor pair, which implies

$$\|\eta_h(t)\|_{L^2} \le c \sup_{\mathbf{v}_h \in \mathbf{V}_h} \frac{b(\mathbf{v}_h, \eta_h(t))}{\|\mathbf{v}_h\|_1}.$$

Using the fundamental relation (2.31) this implies

$$\begin{aligned}\|\eta_h(t)\|_{L^2} &\le c \sup_{\mathbf{v}_h \in \mathbf{V}_h} \frac{(\theta_h'(t), \mathbf{v}_h)_{L^2} + a(\theta_h(t), \mathbf{v}_h) + (\rho_h'(t), \mathbf{v}_h)_{L^2}}{\|\mathbf{v}_h\|_1} \\ &\le c\big(\|\theta_h'(t)\|_{L^2} + |\theta_h(t)|_1 + \|\rho_h'(t)\|_{L^2}\big),\end{aligned}$$

and thus using (2.34), (2.35), (2.29) we get

$$\Big(\int_0^t \|\eta_h(\tau)\|_{L^2}^2 \, d\tau\Big)^{\frac{1}{2}} \le c|\theta_h(0)|_1 + c\sqrt{t}h^m E_t(\mathbf{u}, p).$$

Finally, using (2.30) yields

$$\begin{aligned}\Big(\int_0^t \|p(\tau) - p_h(\tau)\|_{L^2}^2 \, d\tau\Big)^{\frac{1}{2}} &\le \Big(\int_0^t \|\eta_h(\tau)\|_{L^2}^2 \, d\tau\Big)^{\frac{1}{2}} + \sqrt{t} \max_{0\le\tau\le t} \|\xi_h(\tau)\|_{L^2} \\ &\le c|\mathbf{e}_h(0)|_1 + c\sqrt{t}h^{m-1}E_t(\mathbf{u}, p),\end{aligned}$$

and thus the pressure error bound (2.24) holds. □

Remark 2.1.15 The error bounds in this theorem show an optimal behavior w.r.t. the rate of convergence for $h \downarrow 0$. If we use the Hood-Taylor pair with index $k \ge 2$ (i.e., degree k polynomials for velocity and degree $k-1$ for pressure) and assume that the solution pair $(\mathbf{u}, p)$ is sufficiently smooth ($m = k+1$) then Theorem 2.1.14 yields an h^{k+1} bound for the velocity L^2-error and an h^k bound both for the velocity H^1-error and the pressure L^2-error. From the result in (2.22) we see that in the L^2-norm the discretization error in the initial condition is exponentially damped for increasing t. The term $E_t(\mathbf{u}, p)$ quantifies the smoothness of the solution pair $(\mathbf{u}, p)$ on $[0, t]$. Furthermore note that in (2.23), (2.24) apart from a constant c (independent of t) and the smoothness measure $E_t(\mathbf{u}, p)$ there are additional terms $h\sqrt{t}$ and $\sqrt{t}$, respectively, which grow as functions of t. Such a t-dependent factor does not occur in (2.22).

The same semi-discretization approach can be applied to the weak formulation of the *Navier*-Stokes problem (1.50). The semi-discrete Navier-Stokes problem is as follows: Find $\mathbf{u}_h(t) \in \mathbf{V}_h$, $p_h(t) \in Q_h$ such that $\mathbf{u}(0) = \mathbf{u}_{0,h}$ and for all $\mathbf{v}_h \in \mathbf{V}_h$ and all $q_h \in Q_h$:

$$\begin{aligned}\frac{d}{dt} m(\mathbf{u}_h(t), \mathbf{v}_h) + a(\mathbf{u}_h(t), \mathbf{v}_h) + c(\mathbf{u}_h(t); \mathbf{u}_h(t), \mathbf{v}_h) + b(\mathbf{v}_h, p_h(t)) &= (\mathbf{g}, \mathbf{v}_h)_{L^2} \\ b(\mathbf{u}_h(t), q_h) &= 0,\end{aligned}$$

for all $t \in [0, T]$. Using the representations as in (2.20) this Galerkin discretization can be rewritten as

Determine $\vec{\mathbf{u}}(t) \in \mathbb{R}^N$, $\vec{\mathbf{p}}(t) \in \mathbb{R}^K$ with with $\vec{\mathbf{u}}(0) = \vec{\mathbf{u}}_0$ and $(p_h(t), 1)_{L^2} = 0$ such that

$$\begin{aligned}\mathbf{M}\frac{d\vec{\mathbf{u}}}{dt}(t) + \mathbf{A}\vec{\mathbf{u}}(t) + \mathbf{N}(\vec{\mathbf{u}}(t))\vec{\mathbf{u}}(t) + \mathbf{B}^T\vec{\mathbf{p}}(t) &= \vec{\mathbf{f}} \\ \mathbf{B}\vec{\mathbf{u}}(t) &= 0,\end{aligned} \tag{2.37}$$

for all $t \in [0, T]$.

The nonlinear operator $\mathbf{N}$ is given by

$$\mathbf{N}(\vec{\mathbf{u}}) = \mathbf{N}(\mathbf{u}_h) \in \mathbb{R}^{N\times N}, \quad \mathbf{N}(\vec{\mathbf{u}})_{ij} = \int_\Omega (\mathbf{u}_h \cdot \nabla \boldsymbol{\xi}_j) \cdot \boldsymbol{\xi}_i \, dx.$$

We obtain a nonlinear system of differential algebraic equations (DAEs) for the unknown vector functions $\vec{\mathbf{u}}(t)$, $\vec{\mathbf{p}}(t)$. For a discretization error analysis of this problem we refer to the literature [25, 26].

2.2 Numerical experiments

After the theoretical analysis in Section 2.1 we now present a numerical study of the convergence behavior of the Hood-Taylor pair for $k = 2$, cf. Theorem 2.1.11. We consider two numerical experiments, namely the flow in a rectangular tube (Section 2.2.1) and the flow in a curved channel (Section 2.2.2).

2.2.1 Flow in a rectangular tube

Let $\Omega = (0, L) \times (0, 1)^2$ be a rectangular tube of length $L > 0$. Consider (2.9) with $\xi = 0, \mathbf{w} = 0$ (i. e., the stationary Stokes case) where we prescribe the boundary conditions $(\mathbf{u}, p)|_{x_1=0} = (\mathbf{u}, p)|_{x_1=L}$ (periodic boundary conditions in x_1-direction) and $\mathbf{u} = 0$ on the remaining boundaries. The right-hand side is set to $\mathbf{g} = (1, 0, 0)$, which can be interpreted as gravity force in x_1-direction. Then the analytic solution is given by $\mathbf{u}(x) = (Re\, s(x_2, x_3), 0, 0)$ and $p = 0$, where s is the solution of the 2D Poisson problem

$$-\Delta s = 1 \qquad \text{on } (0, 1)^2,$$

with homogeneous Dirichlet boundary conditions. By Fourier analysis, s can be expressed in terms of the Fourier series

$$s(x_2, x_3) = \frac{16}{\pi^4} \sum_{i,j=1}^{\infty} \alpha_{2i-1,2j-1} s_{2i-1,2j-1}(x_2, x_3), \quad (x_2, x_3) \in (0, 1)^2, \tag{2.38}$$

with $\alpha_{i,j} = \frac{1}{ij(i^2+j^2)}$ and $s_{i,j}(x_2, x_3) = \sin(i\pi x_2) \sin(j\pi x_3)$. In Figure 2.2 we give a plot of s.

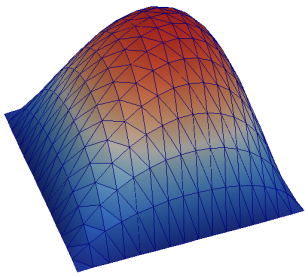

Fig. 2.2. Solution s of the Poisson equation on the unit square.

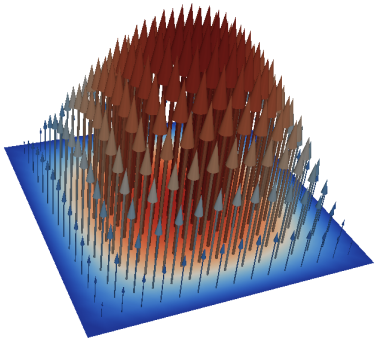

Fig. 2.3. Velocity $\mathbf{u}_h$ visualized on slice $x_1 = L/2$.

The initial triangulation $\mathcal{T}_0$ is constructed by subdividing Ω into $4 \times 1 \times 1$ sub-cubes each consisting of 6 tetrahedra. Then $\mathcal{T}_0$ is successively uniformly refined 5 times by applying the regular refinement rule yielding $\mathcal{T}_1, \ldots, \mathcal{T}_5$. In our experiments we used $L = 4$ and $Re = 1$. In Figure 2.3 the discrete velocity $\mathbf{u}_h$ is illustrated for the triangulation $\mathcal{T}_3$.

The discrete pressure p_h is equal to zero (up to machine accuracy). Table 2.1 shows the dimension of the finite element spaces $\mathbf{V}_h$ and Q_h and the convergence of $\mathbf{u}_h$ to $\mathbf{u}$ w.r.t. the L^2 and H^1 norm for different refinement levels. We observe third order convergence w.r.t. the L^2 norm and second order convergence w.r.t. the H^1 norm. These are the optimal rates which can be expected in view of Theorem 2.1.11 and the L^2 bound (2.13), as $\mathbf{u}$ and p are sufficiently smooth.

Note the (very) high dimension of the velocity space on level 5, due to the fact that the number of tetrahedra grows with a factor of 8 in each refinement. Furthermore it is clear that the dimension of the pressure space is much smaller than that of the velocity space.

# ref.	$\dim \mathbf{V}_h$	$\dim Q_h$	$\|\mathbf{u}-\mathbf{u}_h\|_{L^2}$	order	$\|\mathbf{u}-\mathbf{u}_h\|_1$	order
0	24	16	9.63 E-3	—	7.17 E-2	—
1	432	72	2.06 E-3	2.22	1.89 E-2	1.92
2	4 704	400	2.18 E-4	3.24	4.01 E-3	2.24
3	43 200	2 592	2.49 E-5	3.13	9.40 E-4	2.09
4	369 024	18 496	3.02 E-6	3.04	2.30 E-4	2.03
5	3 048 192	139 392	3.01 E-7	3.33	5.75 E-5	2.00

Table 2.1. Dimension of the finite element spaces and convergence behavior w.r.t. L^2 and H^1 norm for different refinement levels.

2.2.2 Flow in a curved channel

Let $\Omega = \{ x \in \mathbb{R}^3 : 0 < x_1 < L, -a(x_1) < x_2 < a(x_1), 0 < x_3 < 1 \}$ be a channel of length $L > 0$ with $a : [0, L] \to (0, \infty)$ defining the shape in x_2-direction, cf. Figure 2.4. In our experiment we set $L = 4$ and $a(x_1) = e^{-\alpha x_1}$, $\alpha = 1/4$. Consider (2.9) with $\xi = 0, \mathbf{w} = 0$ (i. e., the stationary Stokes case). We take the pair $(\mathbf{u}, p)$ given by

$$u_1(x) = \frac{1 - \left(\frac{x_2}{a(x_1)}\right)^2}{a(x_1)}, \quad u_2(x) = \frac{u_1(x)\, x_2\, a'(x_1)}{a(x_1)}, \quad u_3(x) = 0,$$
$$p = \frac{1}{2} x_2^2 \alpha e^{\alpha x_1}(-\alpha^2 + 6e^{2\alpha x_1}) + \alpha e^{\alpha x_1} - \frac{2}{3\alpha} e^{3\alpha x_1}.$$

The velocity field $\mathbf{u}$ is divergence free. We take the right-hand side $\mathbf{f} := (-\frac{1}{2} x_2^2 \alpha^2 e^{\alpha x_1}(\alpha^2 - 36e^{2\alpha x_1}), -9x_2^3 \alpha^3 e^{3\alpha x_1}, 0)$ such that the pair $(\mathbf{u}, p)$ is a solution of (2.9). In our experiment we use boundary conditions $(\mathbf{u}, p)|_{x_3=0} = (\mathbf{u}, p)|_{x_3=1}$ (periodic boundary conditions in x_3-direction) and Dirichlet boundary conditions for $\mathbf{u}$ on the remaining part of the boundary.

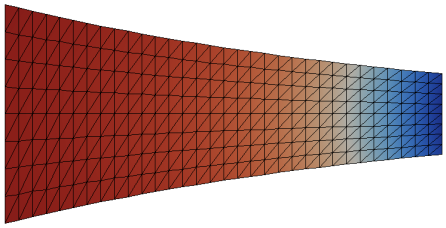

Fig. 2.4. Grid and pressure p_h visualized on slice $x_3 = 0.5$.

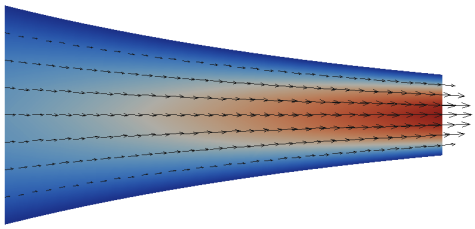

Fig. 2.5. Velocity $\mathbf{u}_h$ visualized on slice $x_3 = 0.5$.

For the discretization of Ω we first introduce the auxiliary domain $\hat{\Omega} = (0, L) \times (-1, 1) \times (0, 1)$, which is discretized by $4 \times 1 \times 1$ subcubes each subdivided into 6 tetrahedra. The resulting initial triangulation $\hat{\mathcal{T}}_0$ is then successively uniformly refined 5 times applying the regular refinement rule yielding $\hat{\mathcal{T}}_1, \ldots, \hat{\mathcal{T}}_5$. Using the mapping $F : \mathbb{R}^3 \to \mathbb{R}^3$, $F(x) = (x_1, a(x_1)x_2, x_3)$ the vertices of $\hat{\mathcal{T}}_\ell$ are mapped to the physical domain Ω. For each level $\ell = 0, \ldots, 5$, this induces corresponding triangulations $\mathcal{T}_\ell$ of polygonal domains Ω_ℓ *approximating* Ω. The mapping is such that all vertices on the boundary of the triangulation lie on the respective boundaries of Ω. At the $x_2 = \pm a(x_1)$ part of the boundary, however, the boundary faces of the triangulation do *not* coincide with the curved boundary. A 2D slice of the triangulation $\mathcal{T}_3$ and the corresponding pressure solution p_h is shown in Figure 2.4. The velocity field $\mathbf{u}_h$ is depicted in Figure 2.5.

Table 2.2 shows the dimension of the finite element spaces $\mathbf{V}_h$ and Q_h and the convergence of $\mathbf{u}_h$ to $\mathbf{u}$ w.r.t. L^2 and H^1 norm for different refinement levels. For small grid sizes the

# ref.	dim $\mathbf{V}_h$	dim Q_h	$\|\mathbf{u}-\mathbf{u}_h\|_{L^2}$	order	$\|\mathbf{u}-\mathbf{u}_h\|_1$	order	$\|p-p_h\|_{L^2}$	order
0	42	10	1.45 E-1	—	8.39 E-1	—	3.09 E+0	—
1	540	54	9.76 E-3	3.89	9.02 E-2	3.22	6.18 E-1	2.74
2	5 208	340	9.50 E-4	3.36	1.43 E-2	2.66	1.48 E-1	2.05
3	45 360	2 376	2.19 E-4	2.12	3.43 E-3	2.06	3.65 E-2	2.02
4	377 952	17 680	5.96 E-5	1.88	1.15 E-3	1.58	9.07 E-3	2.01
5	3 084 480	136 224	1.57 E-5	1.92	4.11 E-4	1.48	2.27 E-3	2.00

Table 2.2. Dimension of the finite element spaces and convergence behavior w.r.t. L^2 and H^1 norm for different refinement levels.

convergence order of the velocity error w.r.t. the H^1 norm tends to 1.5. This suboptimal behavior is due to the fact that a part of $\partial\Omega$ is curved and is approximated by a piecewise polygonal approximation. It is known, cf. [38], that due to this boundary approximation, with quadratic finite elements in general one has only a suboptimal $\mathcal{O}(h^{1.5})$ error behavior (w.r.t. the H^1 norm). The optimal order of 2 (cf. experiment in the previous section) can be achieved by using so-called *isoparametric* elements which are defined on curved tetrahedra by applying a non-affine transformation to the reference tetrahedron. A short discussion on this topic can be found in Section 2.3.

2.3 Discussion and additional references

In these lecture notes we restrict to *finite element* discretization approaches for the incompressible Navier-Stokes equations. Other important discretization methods, which are particularly popular among engineers, are based on *finite volume* techniques. For an introduction to these methods we refer to [44, 18, 43].

We treated only the (very popular) Hood-Taylor finite element spaces. These spaces are members of the family of *conforming* finite element spaces, which means that the spaces $(\mathbf{V}_h, Q_h)$ used for discretization are subspaces of the spaces in which the weak formulation is well-posed. In our setting we have $\mathbf{V}_h \subset H_0^1(\Omega)^3$, $Q_h \subset L_0^2(\Omega)$. Below we briefly address a few issues related to other finite element techniques for (Navier-)Stokes equations.

Other LBB stable pairs of conforming spaces. There are other conforming finite element spaces that are used for the discretization of (Navier-)Stokes equations. We mention two well-known techniques. For this we need the barycentric coordinates defined in Definition 2.1.2. Let $b_T(x) := \Pi_{i=1}^4 \lambda_i(x)$ be the product of the barycentric coordinates for $x \in T$. This "bubble function" is a polynomial of degree 4 in T that is zero on ∂T. It is extended by zero values outside T. Define

$$B_4 := \left\{ v \in C(\overline{\Omega}) : v_{|T} \in \mathrm{span}(b_T) \quad \text{for all} \quad T \in \mathcal{T}_h \right\}$$

The *"mini-element"* is defined as the pair of spaces $(\mathbf{V}_h, Q_h)$ with

$$\mathbf{V}_h = (\mathbb{X}_{h,0}^1 \oplus B_4)^3, \quad Q_h = \mathbb{X}_h^1 \cap L_0^2(\Omega),$$

i.e., for the velocity we use the space of continuous piecewise linears extended by the space of bubble functions, and for the pressure we use the space of continuous piecewise linears. An advantage of this pair compared to the $P_2 - P_1$ Hood-Taylor pair is that the mini-element allows a simpler data structure: for the bubble functions one has one unknown per velocity component in each tetrahedron and all other unknowns for velocity and pressure are located at the vertices of the tetrahedra. The mini-element is one order less accurate than the $P_2 - P_1$ Hood-Taylor pair.

In the *Crouzeix-Raviart* pair of spaces one uses *discontinuous pressure* approximations. For the velocity one uses continuous piecewise polynomials, enriched (in order to guarantee stability) with bubble functions. More precisely:

$$\mathbf{V}_h = \{\, \mathbf{v} \in C(\overline{\Omega})^3 : \mathbf{v}_{|T} \in (\mathbb{X}^2_{h,0} \oplus B_4)^3 \quad \text{for all} \quad T \in \mathcal{T}_h \,\}$$
$$Q_h = \{\, q \in L^2_0(\Omega) : q_{|T} \in \mathcal{P}_1 \quad \text{for all} \quad T \in \mathcal{T}_h \,\}.$$

Similar Crouzeix-Raviart pairs are defined using higher order polynomials. Both the mini-element and the Crouzeix-Raviart pair are LBB-stable and conforming spaces. The enrichment of the velocity space using the bubble functions is important for the LBB stability property to hold. An extensive treatment of these spaces is given in e.g. [20, 34].

LBB stable nonconforming spaces. In the nonconforming case one uses a finite element pair $(\mathbf{V}_h, Q_h)$ with $\mathbf{V}_h \nsubseteq \mathbf{V}$ or $Q_h \nsubseteq Q$. An example from this class of finite element methods is the lowest order *nonconforming Crouzeix-Raviart* pair. We explain this pair. For a tetrahedron T the set of its 4 faces is denoted $\mathcal{F} = \{F\}$. The barycenter (center of gravity) of a triangle F is denoted by C_F. We introduce the space of piecewise linear functions:

$$\mathbf{V}_h^{CR} := \{\, \mathbf{v}_h \in L^2(\Omega)^3 \mid \forall\, T \in \mathcal{T}_h : \ (\mathbf{v}_h)_{|T} \in \mathcal{P}_1, \quad [\mathbf{v}_h]_F(C_F) = 0 \quad \forall F \in \mathcal{F}, \\ \mathbf{v}_h(C_F) = 0 \quad \forall F \subset \partial\Omega \,\}.$$

Here $[\mathbf{v}_h]_F$ denotes the jump of $\mathbf{v}_h$ across the face F. Due to the fact that functions from $\mathbf{V}_h^{CR}$ are not necessarily continuous across the faces of a tetrahedron, this space is nonconforming: $\mathbf{V}_h^{CR} \nsubseteq \mathbf{V}$. If for the pressure one uses the (conforming) space $\mathbb{X}_h^0$ of piecewise constants, then the pair $(\mathbf{V}_h^{CR}, \mathbb{X}_h^0)$ is LBB stable. A detailed treatment of this and other nonconforming pairs is given in [8, 12].

Unstable pairs: stabilization. Instead of using an LBB stable pair of finite element spaces for discretization of a saddle point problem, one can also use an unstable pair and apply the technique of stabilization. We outline a popular technique introduced in [28]. Let $(\mathbf{V}_h, Q_h)$ be a pair of conforming finite element spaces, not necessarily LBB stable. To simplify the presentation we assume that the pressure space Q_h contains only continuous pressure functions. We consider the stationary Stokes problem in weak formulation: determine $(\mathbf{u}, p) \in \mathbf{V} \times Q = H^1_0(\Omega)^3 \times L^2_0(\Omega)$ such that

$$a(\mathbf{u}, \mathbf{v}) + b(\mathbf{v}, p) - b(\mathbf{u}, q) = (\mathbf{g}, \mathbf{v})_{L^2} \quad \text{for all} \quad (\mathbf{v}, q) \in \mathbf{V} \times Q. \tag{2.39}$$

The stabilized discretization reads as follows: determine $(\mathbf{u}_h, p_h) \in \mathbf{V}_h \times Q_h$ such that

$$a(\mathbf{u}_h, \mathbf{v}_h) + b(\mathbf{v}_h, p_h) - b(\mathbf{u}_h, q_h) + \delta s_h(\mathbf{u}_h, p_h; \mathbf{v}_h, q_h) = (\mathbf{g}, \mathbf{v}_h)_{L^2} + \delta g_h(\mathbf{v}_h, q_h)$$

for all $(\mathbf{v}_h, q_h) \in \mathbf{V}_h \times Q_h$, with the stabilization terms

$$s_h(\mathbf{u}_h, p_h; \mathbf{v}_h, q_h) \\ := \sum_{T \in \mathcal{T}_h} h_T^3 \int_T \big(-\frac{1}{Re}\Delta \mathbf{u}_h + \nabla p_h \big) \cdot \big(-\frac{1}{Re}\Delta \mathbf{v}_h + \nabla q_h \big) + \operatorname{div} \mathbf{u}_h \operatorname{div} \mathbf{v}_h \, dx,$$
$$g_h(\mathbf{v}_h, q_h) := \sum_{T \in \mathcal{T}_h} h_T^3 \int_T \mathbf{g} \cdot \big(-\frac{1}{Re}\Delta \mathbf{v}_h + \nabla q_h \big) \, dx,$$

and a stabilization parameter $\delta > 0$. Due to the stabilization term it can be shown that the bilinear form

$$(\mathbf{u}_h, p_h; \mathbf{v}_h, q_h) \to a(\mathbf{u}_h, \mathbf{v}_h) + b(\mathbf{v}_h, p_h) - b(\mathbf{u}_h, q_h) + \delta s_h(\mathbf{u}_h, p_h; \mathbf{v}_h, q_h)$$

satisfies a discrete inf-sup condition (in suitable norms) on $\mathbf{V}_h \times Q_h$ with a strictly positive inf-sup constant independent of h. This inf-sup property then leads to (optimal) discretization error bounds. Note that in such a method one has to choose an "appropriate" value for the stabilization parameter δ. In the literature this stabilization technique is known as a *Galerkin/Least-Squares method* (GaLS). To explain this name, we consider the Stokes problem in the formal operator form

$$\begin{pmatrix} A & B^T \\ B & 0 \end{pmatrix} \begin{pmatrix} \mathbf{u} \\ p \end{pmatrix} = \begin{pmatrix} \mathbf{g} \\ 0 \end{pmatrix}.$$

In the stabilization we add a term of the *least-squares* form

$$\Big\langle \begin{pmatrix} A & B^T \\ B & 0 \end{pmatrix} \begin{pmatrix} \mathbf{u}_h \\ p_h \end{pmatrix} - \begin{pmatrix} \mathbf{g} \\ 0 \end{pmatrix}, \begin{pmatrix} A & B^T \\ B & 0 \end{pmatrix} \begin{pmatrix} \mathbf{v}_h \\ q_h \end{pmatrix} \Big\rangle.$$

The discretization method is consistent in the sense that if in the stabilization term $(\mathbf{u}_h, p_h)$ is replaced by the solution $(\mathbf{u}, p)$ of the continuous problem (2.39) then the stabilization term is equal to zero. Due to this one still has the important Galerkin orthogonality property. For more explanation and other stabilization techniques we refer to the literature, e.g. [16, 19, 34, 41].

Spectral and hp-finite element methods In these lecture notes we restrict ourselves to the class of h-finite element methods. This means that in the finite element spaces we use polynomials of a fixed low degree (e.g. the P_2-P_1 Hood-Taylor pair) and a desired accuracy of the discretization is obtained by taking a mesh size that is sufficiently small ("h-refinement"). An alternative is to use a fixed mesh size h and then use polynomials of (very) high degree to obtain an accurate discretization ("p-refinement"). This leads to the class of so-called *spectral methods*, cf. [34]. If one uses a hybrid approach in the sense that in the discretization both the mesh size and the polynomial degree (per element T in the triangulation) are varied this leads to a *hp*-finite element method, cf. [36].

Discontinuous Galerkin techniques. In the past few years the discontinuous Galerkin method (DG) has received much attention. In this approach one uses finite element spaces consisting of piecewise polynomials without inter-element continuity requirement, i.e. instead of the space $\mathbb{X}_h^k$, $k \geq 1$, in (2.2b) one uses

$$\mathbb{X}_{h,DG}^k := \left\{ v : \Omega \to \mathbb{R} : \; v_{|T} \in \mathcal{P}_k \;\; \text{for all} \;\; T \in \mathcal{T}_h \right\}.$$

In order to enforce smoothness across the faces of the elements T, the bilinear form is modified by adding suitable jump terms of the discrete test and trial functions across the element faces. The DG approach can be combined with stabilization techniques and due to the nice locality property (no continuity requirement across the faces in the finite element space) its use in an *hp*-finite element setting is very natural. Discontinuous Galerkin methods turn out to be particularly suitable for hyperbolic and convection-dominated problems. For further information we refer to the literature, e.g. [18, 2, 24].

Isoparametric finite elements. A further issue that is relevant in the context of finite element methods is how to treat *curved boundaries*. The effect of an inaccurate boundary approximation is seen in the numerical experiment in Section 2.2.2. For an accurate boundary approximation so-called isoparamtric finite elements are very useful. For a treatment of this standard finite element technique we refer to the literature, e.g. [11, 6, 5].

Mass conservation property. In an incompressible flow problem in weak formulation the mass conservation property is described by

$$\int_\Omega q \operatorname{div} \mathbf{u} \, dx = 0 \quad \text{for all} \;\; q \in L_0^2(\Omega).$$

Hence, $\operatorname{div} \mathbf{u} = 0$ holds on Ω, in L^2-sense. In the discrete problem, with finite element spaces $\mathbf{V}_h$ and Q_h for velocity and pressure, respectively, one obtains the variational equation

$$\int_\Omega q_h \operatorname{div} \mathbf{u}_h \, dx = 0 \quad \text{for all} \;\; q_h \in Q_h, \tag{2.40}$$

for the discrete velocity solution $\mathbf{u}_h \in \mathbf{V}_h$. Such a function $\mathbf{u}_h$ is also called *discretely divergence-free*. In general this does *not* imply $\operatorname{div} \mathbf{u}_h = 0$ in Ω (in L^2-sense) due to

$Q_h \neq L^2_0(\Omega)$. This leads to the issue of how well mass is conserved in the finite element discretization. We briefly address two approaches that apply to the Hood-Taylor finite element method and result in discretizations with good mass conservation properties.

The first method is from [40] and is based on an extension of the pressure finite element space by piecewise constants. Let $\mathbb{X}^k_h$ and $\mathbb{X}^k_{h,0}$ be the polynomial finite element spaces introduced in section 2.1. For the velocity discretization we use the same space as in the Hood-Taylor method , i.e., $\mathbf{V}_h = (\mathbb{X}^k_{h,0})^d$, $k \geq 2$. For the pressure we extend the Hood-Taylor pressure space by the space of piecewise constants, i.e., we take $Q_h = (\mathbb{X}^{k-1}_h \cup \mathbb{X}^0_h) \cap L^2_0(\Omega)$. In [40] it is proved that for $k = 2$, $d = 2$, the pair $(\mathbf{V}_h, Q_h)$ is LBB-stable. In the discrete variational equation (2.40) corresponding to this finite element pair one can take, for $T \in \mathcal{T}_h$, the function $q_h(x) = 1 + c$ if $x \in T$, $q_h(x) = c$ otherwise, with a constant c such that $\int_\Omega q_h \, dx = 0$ holds. This results in

$$\begin{aligned} 0 &= \int_T \operatorname{div} \mathbf{u}_h \, dx + c \int_\Omega \operatorname{div} \mathbf{u}_h \, dx \\ &= \int_T \operatorname{div} \mathbf{u}_h \, dx + c \int_{\partial\Omega} \mathbf{u}_h \cdot \mathbf{n} \, ds = \int_T \operatorname{div} \mathbf{u}_h \, dx, \end{aligned}$$

where in the last identity we used that $\mathbf{u}_h = 0$ on $\partial\Omega$. Hence we obtain a *local mass-conservation property* in the sense that $\int_T \operatorname{div} \mathbf{u}_h \, dx = 0$ holds for all $T \in \mathcal{T}_h$.

The second method is based on the so-called Scott-Vogelius finite element pair [37]. In this pair, for the velocity one uses the same space as in the Hood-Taylor method, i.e., $\mathbf{V}_h = (\mathbb{X}^k_{h,0})^d$, $k \geq 2$, and for the pressure one uses piecewise polynomials of degree $k-1$ that are not necessarily continuous, i.e.

$$Q_h = \left\{ v \in L^2_0(\Omega) : \, v_{|T} \in \mathcal{P}_{k-1} \quad \text{for all} \;\; T \in \mathcal{T}_h \right\}.$$

In [37] it is proved that for $d = 2$ the pair $(\mathbf{V}_h, Q_h)$ with $k \geq 4$ is LBB-stable provided the mesh does not contain any so-called nearly-singular vertices. Further analyses of this pair can be found in [47, 48, 49]. We outline two stability results for $d = 3$. For this we need special tetrahedral grids. Starting from a regular tetrahedral triangulation $\mathcal{T}_h$ a so-called Hsieh-Clough-Tocher triangulation is obtained by subdiving each tetrahedron $T \in \mathcal{T}_h$ into 4 subtetrahedra by connecting the barycenter of T to the four vertices of T. If each tetrahedron of this Hsieh-Clough-Tocher triangulation is further refined into 3 subtetrahedra by connecting the barycenter of T to each of the four barycenters of the four tetrahedra adjacent to T one obtains a so-called Powell-Sabin triangulation. In [47] it is proved that on Hsieh-Clough-Tocher triangulations the Scott-Vogelius pair $(\mathbf{V}_h, Q_h)$ is LBB-stable for $k \geq 3$. In [49] it is proved that on Powell-Sabin triangulations the pair $(\mathbf{V}_h, Q_h)$ is LBB-stable for $k = 2$. For the Scott-Vogelius pair $(\mathbf{V}_h, Q_h)$ one can take $q_h = \operatorname{div} \mathbf{u}_h + c$ in (2.40), with a constant c such that $\int_\Omega q_h \, dx = 0$ holds. Then one obtains

$$\int_\Omega (\operatorname{div} \mathbf{u}_h)^2 \, dx = 0,$$

and thus the discrete velocity is divergence-free, i.e., a discrete mass conservation property $\operatorname{div} \mathbf{u}_h = 0$ in Ω (in L^2-sense) holds. A relation between the Scott-Vogelius pair and the Hood-Taylor pair is derived in [9]. There it is shown that if the Hood-Taylor discretization is applied to the Navier-Stokes equation with grad-div stabilization, i. e., one adds a (consistent) term $\gamma(\operatorname{div} \mathbf{u}_h, \operatorname{div} \mathbf{v}_h)_{L^2}$ to the disrete momentum equation, then the resulting discrete velocity $\mathbf{u}_h$ tends to $\mathbf{u}_h^{SV}$ if $\gamma \to \infty$. Here $\mathbf{u}_h^{SV}$ denotes the divergence-free discrete velocity solution obtained by using the Scott-Vogelius pair Hence, for the Hood-Taylor pair the mass conservation property can be improved (significantly) if grad-div stabilization is used.

Type of triangulations. In these lecture notes we restrict ourselves to finite elements on *tetrahedral* triangulations. Finite element techniques, however, can also be applied using a subdivision

of the three-dimensional domain into, for example, hexahedra or prisms. One can even use subdivisions consisting of combinations of tetrahedra, hexahedra and prisms. An important class are the tensor product finite elements (in the 2D case, these are also called quadrilateral finite elements). For a treatment of these we refer to standard finite element literature.

3

Time integration of semi-discrete Navier-Stokes equations

3.1 Introduction

Let $I := [0, t_e]$, $f : I \to \mathbb{R}^N$, $F : \mathbb{R}^N \to \mathbb{R}^N$ and $u_0 \in \mathbb{R}^N$. Consider an initial value problem: determine $u(t) \in \mathbb{R}^N$ such that

$$\frac{du}{dt} + F(u) = f(t) \quad \text{for} \quad t \in I, \quad u(0) = u_0. \tag{3.1}$$

As we will see further on, the Stokes- and Navier-Stokes systems of DAEs in (2.21) and (2.37) take this form if one eliminates the pressure variable by restricting to the subspace of (discrete) divergence free velocities. Related to existence and uniqueness of a solution of (3.1) we give a standard result from the literature (Picard-Lindelöf theorem). For $b > 0$ define $G_b := \{\, v \in \mathbb{R}^N : \|v - u_0\| \leq b \,\}$ with $\|\cdot\|$ any given norm on $\mathbb{R}^N$. Assume that for $a > 0$ the function $f : [0, a] \to \mathbb{R}^N$ is continuous and that F satisfies the Lipschitz condition:

$$\|F(v) - F(w)\| \leq L\|v - w\| \quad \text{for all} \quad v, w \in G_b. \tag{3.2}$$

Then the initial value problem in (3.1) *has a unique solution* $u(t)$ for $t \in [0, \alpha]$ with $\alpha := \min\{a, bL^{-1}\}$. In the remainder we assume that f and F satisfy these conditions (for suitable a, b, L) and that $t_e \leq \alpha$, i.e., (3.1) has a unique solution.

We discuss a few discretization methods for the general problem (3.1). In our applications the systems are very stiff and thus we need implicit methods. A classical and still very popular method is the θ-scheme:

$$\frac{u^{n+1} - u^n}{\Delta t} + \theta F(u^{n+1}) + (1-\theta)F(u^n) = \theta f(t_{n+1}) + (1-\theta)f(t_n), \tag{3.3}$$

with $\theta \in [0, 1]$. For $\theta = 1$ this is the *implicit Euler scheme* and for $\theta = \frac{1}{2}$ this method is known as the *Crank-Nicolson method*. Another popular method is the *BDF2 scheme*:

$$\frac{3}{2}u^{n+1} - 2u^n + \frac{1}{2}u^{n-1} + \Delta t F(u^{n+1}) = \Delta t f(t_{n+1}). \tag{3.4}$$

Note that the θ-scheme is a *one*-step method, whereas the BDF2 method is a linear *two*-step scheme. Another method that is used in our applications is the following *fractional-step θ-scheme*. For a given $\theta \in (0, \frac{1}{2})$, the fractional-step θ-scheme is based on a subdivision of each time interval $[n\Delta t, (n+1)\Delta t]$ in three subintervals with endpoints $(n+\theta)\Delta t$, $(n+1-\theta)\Delta t$, $(n+1)\Delta t$. For given u^n the approximations $u^{n+\theta}$, $u^{n+1-\theta}$, u^{n+1} at these endpoints are defined by

$$\frac{u^{n+\theta}-u^n}{\theta\Delta t} + \alpha F(u^{n+\theta}) + (1-\alpha)F(u^n) = f(t_n) \tag{3.5a}$$

$$\frac{u^{n+1-\theta}-u^{n+\theta}}{(1-2\theta)\Delta t} + (1-\alpha)F(u^{n+1-\theta}) + \alpha F(u^{n+\theta}) = f(t_{n+1-\theta}) \tag{3.5b}$$

$$\frac{u^{n+1}-u^{n+1-\theta}}{\theta\Delta t} + \alpha F(u^{n+1}) + (1-\alpha)F(u^{n+1-\theta}) = f(t_{n+1-\theta}). \tag{3.5c}$$

Standard measures for the quality of discretization methods for (stiff) initial value problems are consistency, stability, a smoothing property and the amount of dissipativity. Below we treat these quality measures for the methods that we consider.

Consistency

The implicit Euler method has a consistency order of 1. The Crank-Nicolson, BDF2 and fractional-step θ-scheme, with $\theta = 1 \pm \frac{1}{2}\sqrt{2}$, all have consistency order 2. We derive this consistency result for the fractional-step θ-scheme.

Lemma 3.1.1 *Assume an arbitrary $f \in C^2([0,t_e])$ and $\lambda \in \mathbb{R}$. Let $u(t)$ be the solution of $\frac{du}{dt} - \lambda u = f$, $u(0) = u_0$. Let u^{n+1} be the result of the fractional-step θ-scheme* (3.5) *applied to this problem with $u^n := u(t_n)$. Then for $\theta = 1 \pm \frac{1}{2}\sqrt{2}$ we have*

$$|u(t_{n+1}) - u^{n+1}| \le c\,(\Delta t)^3 \tag{3.6}$$

with a constant c independent of Δt and n.

Proof. We take $\theta = 1 \pm \frac{1}{2}\sqrt{2}$. For the solution $u(t)$ we have

$$u(t) = e^{\lambda(t-t_n)}u(t_n) + \int_{t_n}^{t} e^{\lambda(t-\tau)} f(\tau)\,d\tau, \quad t \ge t_n.$$

Hence, with $z := \lambda\Delta t$,

$$u(t_{n+1}) = e^{z}u(t_n) + \int_{t_n}^{t_{n+1}} e^{\lambda(t_{n+1}-\tau)} f(\tau)\,d\tau\ .$$

A straightforward calculation, in which we use that $2\theta^2 - 4\theta + 1 = 0$ holds, results in

$$\begin{aligned} u^{n+1} &= g(z)u^n + \Delta t\big[\theta(1+z) - (1-\alpha)\theta^2 z + \mathcal{O}(z^2)\big] f(t_n) \\ &+ \Delta t\big[(1-\theta)(1+\theta z) + (1-\alpha)(3\theta^2 - 4\theta + 1)z + \mathcal{O}(z^2)\big] f(t_{n+1-\theta}) \\ &= g(z)u^n + \Delta t\big(\theta(1+z)f(t_n) + (1-\theta)(1+\theta z)f(t_{n+1-\theta})\big) + \mathcal{O}(\Delta t^3), \end{aligned} \tag{3.7}$$

with

$$g(z) := \frac{\big(1+(1-\alpha)\theta z\big)^2\big(1+\alpha(1-2\theta)z\big)}{\big(1-\alpha\theta z\big)^2\big(1-(1-\alpha)(1-2\theta)z\big)}. \tag{3.8}$$

Taylor expansion results in

$$g(z) = 1 + z + \frac{1}{2}z^2\big[1 + (1-2\alpha)(2\theta^2 - 4\theta + 1)\big] + \mathcal{O}(z^3) \quad (z \to 0).$$

For $\theta = 1 \pm \frac{1}{2}\sqrt{2}$ we have $2\theta^2 - 4\theta + 1 = 0$ and thus

$$g(z) = e^z + \mathcal{O}(z^3) \tag{3.9}$$

holds. The quadrature rule $\int_0^1 v(t)\,dt \approx \xi v(0) + (1-\xi)v(1-\xi)$ is exact for all linear functions v iff $\xi = 1 \pm \frac{1}{2}\sqrt{2}$. Thus for $\theta = 1 \pm \frac{1}{2}\sqrt{2}$ we have

$$\begin{aligned}\int_{t_n}^{t_{n+1}} e^{\lambda(t_{n+1}-\tau)} f(\tau)\,d\tau &= \Delta t\big(\theta e^z f(t_n) + (1-\theta)e^{\theta z} f(t_{n+1-\theta})\big) + \mathcal{O}(\Delta t^3)\\ &= \Delta t\big(\theta(1+z) f(t_n) + (1-\theta)(1+\theta z) f(t_{n+1-\theta})\big) + \mathcal{O}(\Delta t^3).\end{aligned}$$

Using this in combination with (3.7), (3.9) we get

$$\begin{aligned}u^{n+1} &= e^z u(t_n) + \Delta t\big(\theta(1+z) f(t_n) + (1-\theta)(1+\theta z) f(t_{n+1-\theta})\big) + \mathcal{O}(\Delta t^3)\\ &= e^z u(t_n) + \int_{t_n}^{t_{n+1}} e^{\lambda(t_{n+1}-\tau)} f(\tau)\,d\tau + \mathcal{O}(\Delta t^3)\\ &= u(t_{n+1}) + \mathcal{O}(\Delta t^3),\end{aligned}$$

and thus the result is proved. □

A similar bound as in (3.6) can be derived for the case that F is a nonlinear function which satisfies the Lipschitz condition in (3.2). Thus for $\theta = 1 \pm \frac{1}{2}\sqrt{2}$ the fractional-step θ-scheme has consistency order 2.

Remark 3.1.2 For the fractional-step θ-scheme to have consistency order 2 it is *necessary* to take the value $\theta = 1 \pm \frac{1}{2}\sqrt{2}$ in the following sense. Consider the special case $\lambda = 0$, $u_0 = 0$, $f(t) = t$, $n = 0$. Then we have $u(t_1) = u(\Delta t) = \frac{1}{2}(\Delta t)^2$ and a simple computation yields $u^1 = (1-\theta)^2(\Delta t)^2$. Thus for (3.6) to hold *with a θ-value independent of Δt* we need $(1-\theta)^2 = \frac{1}{2}$, i.e., $\theta = 1 \pm \frac{1}{2}\sqrt{2}$.

Remark 3.1.3 Consider the following variant of the fractional-step θ-scheme, with $G(u,t) := F(u) - f(t)$:

$$\begin{aligned}\frac{u^{n+\theta} - u^n}{\theta\Delta t} + \alpha G(u^{n+\theta}, t_{n+\theta}) + (1-\alpha)G(u^n, t_n) &= 0\\ \frac{u^{n+1-\theta} - u^{n+\theta}}{(1-2\theta)\Delta t} + (1-\alpha)G(u^{n+1-\theta}, t_{n+1-\theta}) + \alpha G(u^{n+\theta}, t_{n+\theta}) &= 0\\ \frac{u^{n+1} - u^{n+1-\theta}}{\theta\Delta t} + \alpha G(u^{n+1}, t_{n+1}) + (1-\alpha)G(u^{n+1-\theta}, t_{n+1-\theta}) &= 0.\end{aligned}$$

This scheme is equal to three steps of the θ-scheme (3.3), where α or $1-\alpha$ takes the role of θ in (3.3), and for the three substeps we use time steps $\theta\Delta t$, $(1-2\theta)\Delta t$ and $\theta\Delta t$, respectively. For an accuracy analysis we consider the same test problem as in Lemma 3.1.1 and take $u^n := u(t_n)$, $\theta = 1 \pm \frac{1}{2}\sqrt{2}$. Along the same lines as in the proof of Lemma 3.1.1 one can derive the following, with $z := \lambda\Delta t$:

$$\begin{aligned}u^{n+1} = g(z)u^n &+ (1-\alpha)\Delta t\big[\theta(1+z) f(t_n) + (1-\theta)(1+\theta z) f(t_{n+1-\theta})\big]\\ &+ \alpha\Delta t\big[(1-\theta)\big(1+(1-\theta)z\big) f(t_{n+\theta}) + \theta f(t_{n+1})\big] + \mathcal{O}(\Delta t^3).\end{aligned}$$

For $\int_0^1 v(t)\,dt$ the quadrature rules $\theta v(0) + (1-\theta)v(1-\theta)$ and $(1-\theta)v(\theta) + \theta v(1)$ are exact for all linear functions. Hence, using the Taylor expansion $e^z = 1 + z + \mathcal{O}(z^2)$ we get

$$\begin{aligned}&\int_{t_n}^{t_{n+1}} e^{\lambda(t_{n+1}-\tau)} f(\tau)\,d\tau\\ &= (1-\alpha)\Delta t\big[\theta(1+z) f(t_n) + (1-\theta)(1+\theta z) f(t_{n+1-\theta})\big]\\ &+ \alpha\Delta t\big[(1-\theta)\big(1+(1-\theta)z\big) f(t_{n+\theta}) + \theta f(t_{n+1})\big] + \mathcal{O}(\Delta t^3).\end{aligned}$$

Thus as in the proof of Lemma 3.1.1 we obtain

$$u^{n+1} = e^z u(t_n) + \int_{t_n}^{t_{n+1}} e^{\lambda(t_{n+1}-\tau)} f(\tau)\, d\tau + \mathcal{O}(\Delta t^3) = u(t_{n+1}) + \mathcal{O}(\Delta t^3),$$

Hence, this variant has consistency order 2, too.

Stability

For an error analysis of time discretization methods for stiff problems *stability* properties have to be considered. For a stability analysis, these methods are applied to the test problem

$$\frac{du}{dt} = \lambda u, \qquad \lambda \in \mathbb{C},\ \mathrm{Re}(\lambda) \le 0. \tag{3.10}$$

A solution of this test problem satisfies the growth relation

$$|u(t_{n+1})| = |e^{\lambda \Delta t}||u(t_n)| = e^{\mathrm{Re}(\lambda)\Delta t}|u(t_n)|. \tag{3.11}$$

Due to $\mathrm{Re}(\lambda) \le 0$ the growth factor satisfies $0 \le e^{\mathrm{Re}(\lambda)\Delta t} \le 1$. For one-step methods applied to this test problem one obtains $|u^{n+1}| = g(\lambda\Delta t)|u^n|$, with a so-called *stability function* $g(z)$ which is an approximation of the growth factor $|e^z|$ in (3.11). For the implicit Euler, Crank-Nicolson and fractional-step θ-scheme (cf. (3.8)) the stability function is given by:

$$\begin{aligned}
g_{EB}(z) &:= \left| \frac{1}{1-z} \right| \\
g_{CN}(z) &:= \left| \frac{1+\frac{1}{2}z}{1-\frac{1}{2}z} \right| \\
g_{FS}(z) &:= \left| \frac{\big(1+(1-\alpha)\theta z\big)^2 \big(1+\alpha(1-2\theta)z\big)}{\big(1-\alpha\theta z\big)^2 \big(1-(1-\alpha)(1-2\theta)z\big)} \right|.
\end{aligned}$$

The variant of the fractional-step θ-scheme discussed in Remark 3.1.3 also has the stability function g_{FS}. For the BDF2 method one obtains $u^n = c_0 \left(\frac{2+\sqrt{1+2z}}{3-2z}\right)^n + c_1 \left(\frac{2-\sqrt{1+2z}}{3-2z}\right)^n$, with $z := \lambda\Delta t$ and constants c_0, c_1 that depend on the starting values u^0, u^1. The stability function of the BDF2 method is given by

$$g_{BDF}(z) := \max\left\{ \left| \frac{2+\sqrt{1+2z}}{3-2z} \right|, \left| \frac{2-\sqrt{1+2z}}{3-2z} \right| \right\}.$$

For a given method with stability function g the so-called *stability region* is defined by

$$S := \{\, z \in \mathbb{C} :\ g(z) \le 1 \,\}.$$

The method is said to be A-stable if

$$\mathbb{C}^- := \{\, z \in \mathbb{C} :\ \mathrm{Re}(z) \le 0 \,\} \subset S$$

holds. From standard literature on time discretization methods for (stiff) initial value problems, cf. [23], it is known that the backward -Euler, Crank-Nicolson method and BDF2 method are A-stable.

We consider the fractional-step θ-scheme with $\theta = 1 \pm \frac{1}{2}\sqrt{2}$. Due to the structure of the fractional-step θ-scheme it is natural to restrict to $\alpha \in [0,1]$. First the case $\theta = 1 - \frac{1}{2}\sqrt{2}$ is treated.

Lemma 3.1.4 *Take $\alpha \in [0,1]$, $\theta = 1 - \frac{1}{2}\sqrt{2}$. The fractional-step θ-scheme is A-stable iff $\alpha \in [\frac{1}{2}, 1]$.*

Proof. For $\alpha > 0$ we have

$$\lim_{z\to-\infty} g_{FS}(z) = \left|\frac{1-\alpha}{\alpha}\right| = \frac{1-\alpha}{\alpha}.$$

Since $\frac{1-\alpha}{\alpha} > 1$ for $\alpha < \frac{1}{2}$ the method is not A-stable for $\alpha < \frac{1}{2}$. We consider $\alpha \geq \frac{1}{2}$. The denominator in the function g_{FS} has no zero in $\mathbb{C}^-$ and thus g_{FS} is the norm of a function that is analytic on $\mathbb{C}^-$. From the maximum principle for analytic functions it follows that

$$\max_{z\in\mathbb{C}^-} g_{FS}(z) = \max_{y\in\mathbb{R}} g_{FS}(iy).$$

Due to $g_{FS}(iy) = g_{FS}(-iy)$ we can restrict to $y \in [0,\infty)$. Note that $g_{FS}(0) = 1$ and $\lim_{y\to\infty} g_{FS}(iy) = \frac{1-\alpha}{\alpha} \leq 1$. A straightforward computation yields that on $[0,\infty)$ the derivative of the function $y \to g_{FS}(iy)$ is less than or equal to 0. Hence $\max_{y\in\mathbb{R}} g_{FS}(iy) \leq g_{FS}(0) = 1$. □

We now consider $\theta = 1 + \frac{1}{2}\sqrt{2}$, $\alpha \in [0,1]$. For $\alpha < 1$ the denominator has a zero at $z_0 = (1-\alpha)^{-1}(1-2\theta)^{-1} < 0$. For the value $z = z_0$ the nominator is not equal to zero. Hence $\lim_{z\to z_0} g_{FS}(z) = \infty$ and thus the method is not A-stable. For $\alpha = 1$ it can be shown with the same arguments as in the proof of Lemma 3.1.4 that the method is A-stable. *Below, for the fractional-step θ-scheme we restrict to* $\theta = 1 - \frac{1}{2}\sqrt{2}$, $\alpha \in [\frac{1}{2}, 1]$, or $\theta = 1 + \frac{1}{2}\sqrt{2}$, $\alpha = 1$. For these parameter values the method has consistency order 2 and is A-stable.

Smoothing property

A further criterion which is relevant for comparing these methods is the notion of smoothing, which quantifies the amount of damping of the numerical solutions of (3.10) with λ such that $\mathrm{Re}(\lambda) \to -\infty$ (i.e. of "high" frequencies). For $\mathrm{Re}(\lambda) \to -\infty$ the growth factor $e^{\mathrm{Re}(\lambda)\Delta t}$, cf. (3.11), tends to zero. The smoothing property measures how well this strong damping behavior for $\mathrm{Re}(\lambda) \to -\infty$ is reflected in the numerical scheme. The method has a *smoothing property* if there exists a constant $\delta < 1$ such that for the corresponding stability function g we have

$$\lim_{\mathrm{Re}(z)\to-\infty} g(z) \leq \delta. \tag{3.12}$$

The size of δ is a measure for the strength of the smoothing: a small δ value corresponds to a strong smoothing. A strong smoothing is a desirable property of a numerical scheme. One easily verifies that for the backward Euler and the BDF2 methods we have a maximal smoothing effect, namely with $\delta = 0$. For the Crank-Nicolson method there is no smoothing at all: $\delta = 1$. For the fractional-step θ-method we have a smoothing effect with $\delta = \frac{1}{\alpha} - 1$, and thus the smoothing effect increases for larger α.

Dissipativity

The last property that we consider is the *amount of dissipativity* of a method. This is a measure for the quality of the numerical method when applied to (3.10) with a *periodic* solution of the form $u(t) = e^{ixt}$, $x \in \mathbb{R}$, i.e. with $\lambda = ix$. Hence, in (3.11) we then have a growth factor $e^{\mathrm{Re}(\lambda)\Delta t} = 1$. In this case we have to consider the corresponding stability functions $g(z)$ with $z = ix$, $x \in \mathbb{R}$. The amount of dissipativity is measured by the *deviation* of

$$d(x) := g(ix), \quad x \in \mathbb{R}, \tag{3.13}$$

from the optimal value 1. For the implicit Euler method we have

$$d_{EB}(x) = \frac{1}{\sqrt{1+x^2}},$$

and thus an increasing amount of dissipativity for larger x values. For the Crank-Nicolson we have

$$d_{CN}(x) = 1,$$

and thus *no dissipativity*. For the fractional-step θ-scheme the following holds. For $\theta = 1+\frac{1}{2}\sqrt{2}$, $\alpha = 1$ we have $d_{FS}(x) = g_{FS}(ix) = \left(1 + (1-2\theta)x^2\right)^{\frac{1}{2}}\left(1+\theta^2x^2\right)^{-1}$, which is monotonically decreasing with value 0 for $x \to \infty$, thus in this case there is a large amount of dissipativity for large x values. For the case $\theta = 1 - \frac{1}{2}\sqrt{2}$, $\alpha \in [\frac{1}{2}, 1]$ the dissipativity function depends on α: $d_{FS,\alpha}(x) := g_{FS}(ix)$. We have $\lim_{x\to\infty} d_{FS,\alpha}(x) = \frac{1}{\alpha} - 1$. Inspection of the function $d_{FS,\alpha}$ yields that it is constant for $\alpha = \frac{1}{2}$ and strictly decreasing for $\alpha \in (\frac{1}{2}, 1]$. Furthermore, we have $d_{FS,\alpha'}(x) < d_{FS,\alpha}(x)$ if $\frac{1}{2} \le \alpha < \alpha' \le 1$ and $x > 0$. Thus we have more dissipativity for larger values of α. For a few cases the dissipativity function $d_{FS,\alpha}$ is illustrated in Figure 3.1. Due to $d_{FS}(x) = d_{FS}(-x)$ it suffices to show results for $x \ge 0$.

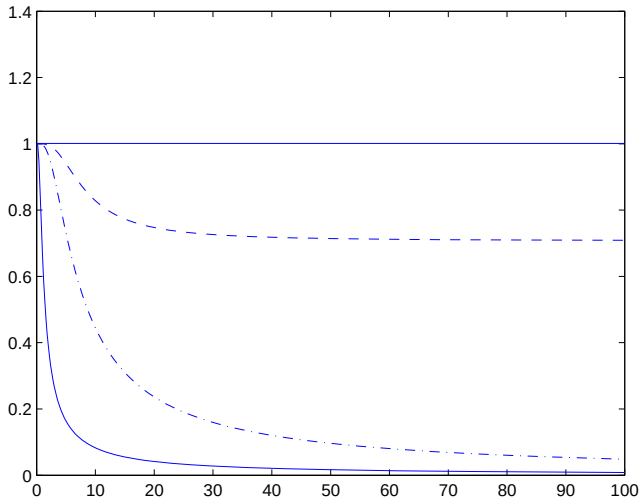

Fig. 3.1. Dissipativity functions $d_{FS,\alpha}$ for $\theta = 1 - \frac{1}{2}\sqrt{2}$, $\alpha \in \{\frac{1}{2}, 2-\sqrt{2}, 1\}$, and $\theta = 1 + \frac{1}{2}\sqrt{2}$, $\alpha = 1$ (top to bottom).

Due to the fact that the stability function g_{FS} is the norm of a rational function we have

$$\lim_{\mathrm{Re}(z)\to-\infty} g_{FS}(z) = \lim_{x\to\infty} g_{FS}(ix).$$

This property also holds for the stability functions of the other three methods. Thus there is a conflict between good smoothing ($\lim_{\mathrm{Re}(z)\to-\infty} g_{FS}(z)$ close to zero, cf. (3.12)) and low dissipativity ($g_{FS}(ix) \approx 1$ for a large range of x values). From the analysis above and Figure 3.1 we see that for $\theta = 1 - \frac{1}{2}\sqrt{2}$ and $\alpha = \frac{1}{2}$ the fractional-step θ-scheme has the same properties as the Crank-Nicolson method, namely no smoothing ($\delta = 1$) and no dissipativity. For $\theta = 1 + \frac{1}{2}\sqrt{2}$, $\alpha = 1$, the fractional-step θ-scheme has properties similar to those of the implicit Euler method: optimal smoothing ($\delta = 0$) and strong dissipativity. A good compromise is found by taking $\theta = 1 - \frac{1}{2}\sqrt{2}$ and $\alpha \in (0, \frac{1}{2})$. A popular parameter choice, cf. [35, 42], is

$$\theta := 1 - \frac{1}{2}\sqrt{2}, \quad \alpha := \frac{1-2\theta}{1-\theta} = 2 - \sqrt{2}. \tag{3.14}$$

For these values (cf. Figure 3.1) the method has "modest" dissipativity and it has a "reasonable" smoothing property with $\delta = \frac{1}{\alpha} - 1 = \frac{1}{2}\sqrt{2}$. Furthermore, due to $\theta\alpha = (1-2\theta)(1-\alpha)$

the systems in (3.5) for the unknowns $u^{n+\theta}$, $u^{n+1-\theta}$, u^{n+1}, respectively, have the same form. *In the remainder we only consider the fractional-step θ-scheme with the parameter values as in* (3.14).

The dissipativity functions $d(x)$ for the implicit Euler, Crank-Nicolson, BDF2 and fractional-step θ-scheme ($\theta = 1 - \frac{1}{2}\sqrt{2}$, $\alpha = 2 - \sqrt{2}$) are illustrated in Figure 3.2. In all four cases we have $d(-x) = d(x)$ and therefore we show the functions only for $x \geq 0$.

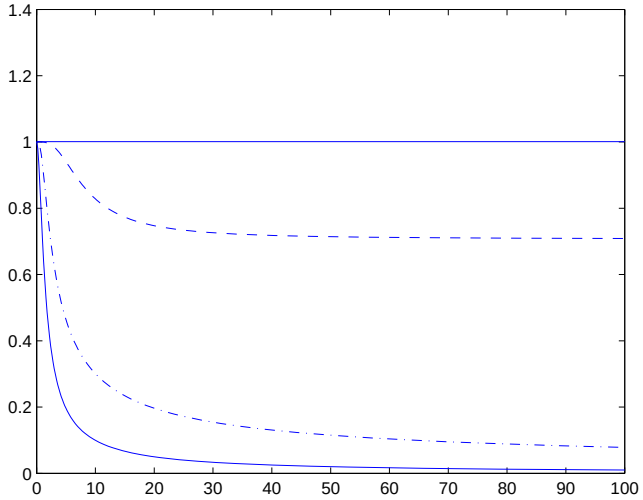

Fig. 3.2. Dissipativity functions d_{CN}, d_{FS}, d_{BDF} and d_{EB} (top to bottom).

In practice often the Crank-Nicolson method is used. A disadvantage of this method, however, is that it has no smoothing property. The fractional-step θ-schema is a method which has both a good smoothing property and modest dissipativity.

In our applications we will use the implicit Euler method (a simple method with a strong smoothing property), the Crank-Nicolson method and the fractional-step θ-scheme. Note that the implicit Euler and the Crank-Nicolson method are special cases of the θ-scheme.

3.2 The θ-scheme for the Navier-Stokes problem

The DAE system (2.37) is rewritten in the form

$$\begin{aligned} \frac{d\vec{\mathbf{u}}}{dt}(t) + \mathbf{M}^{-1}\mathbf{B}^T\vec{\mathbf{p}}(t) &= \mathbf{M}^{-1}\, g(\vec{\mathbf{u}}, t), \\ \mathbf{B}\,\vec{\mathbf{u}}(t) &= 0, \end{aligned} \tag{3.15}$$

where

$$g(\vec{\mathbf{u}}, t) := \vec{\mathbf{f}} - \mathbf{A}\vec{\mathbf{u}}(t) - \mathbf{N}(\vec{\mathbf{u}}(t))\vec{\mathbf{u}}(t).$$

The Stokes DAE system (2.21) has a similar form, with $g(\vec{\mathbf{u}}, t) := \vec{\mathbf{f}} - \mathbf{A}\vec{\mathbf{u}}(t)$. We eliminate the incompressibility constraint $\mathbf{B}\vec{\mathbf{u}}(t) = 0$ and the corresponding Lagrange multiplier $\vec{\mathbf{p}}(t)$ to replace the DAE system by an equivalent ODE system. This can be achieved by applying the $\mathbf{M}$-orthogonal projection $\mathbf{P}$ on $\ker \mathbf{B}$:

$$\mathbf{P} = \mathbf{I} - \mathbf{M}^{-1}\mathbf{B}^T(\mathbf{B}\mathbf{M}^{-1}\mathbf{B}^T)^{-1}\mathbf{B}.$$

The projection $\mathbf{P}$ is orthogonal w.r.t. the scalar product $\langle \cdot, \cdot \rangle_{\mathbf{M}} := \langle \mathbf{M} \cdot, \cdot \rangle$, and $\mathbf{P}\,\vec{\mathbf{v}} = \vec{\mathbf{v}}$ for all $\vec{\mathbf{v}} \in \ker \mathbf{B}$, furthermore $\mathbf{P}\mathbf{M}^{-1}\mathbf{B}^T = 0$. Hence, instead of a DAE system we obtain a system of ordinary differential equations:

A solution $\vec{\mathbf{u}}(t)$ of (3.15) satisfies

$$\frac{d\vec{\mathbf{u}}}{dt}(t) = \mathbf{P}\mathbf{M}^{-1}\, g(\vec{\mathbf{u}}, t). \tag{3.16}$$

If for a given initial condition $\vec{\mathbf{u}}(0) = \vec{\mathbf{u}}_0$, with $\mathbf{B}\mathbf{u}_0 = 0$, and $t \in [0, t_e]$ (with t_e sufficiently small) the problem in (3.16) has a unique solution, then this $\vec{\mathbf{u}}$ is also a solution of (3.15). For a given velocity $\vec{\mathbf{u}}(t)$ the corresponding pressure $\vec{\mathbf{p}}$ is defined by the equation

$$\mathbf{B}\mathbf{M}^{-1}\mathbf{B}^T\vec{\mathbf{p}}(t) = \mathbf{B}\mathbf{M}^{-1}\, g(\vec{\mathbf{u}}, t). \tag{3.17}$$

The matrix $\mathbf{B}\mathbf{M}^{-1}\mathbf{B}^T$ is nonsingular (on the subspace of FE pressure functions with $(p_h, 1)_{L^2} = 0$) due to the LBB stability of the pair of finite element spaces used.

Remark 3.2.1 For the Stokes case we have that $\vec{\mathbf{u}} \to g(\vec{\mathbf{u}}, t)$ is affine and thus $\vec{\mathbf{u}} \to \mathbf{B}\mathbf{M}^{-1}\, g(\vec{\mathbf{u}}, t)$ is affine, too. Hence a Lipschitz condition as in (3.2) is satisfied with a constant L independent of the radius b of the ball G_b. From the Picard-Lindelöf theorem it then follows that for given $\vec{\mathbf{u}}(0) = \vec{\mathbf{u}}_0$ the ODE system (3.16) corresponding to the Stokes problem has a unique solution for $t \in [0, t_e]$ and t_e arbitrary. For the Navier-Stokes case a Lipschitz condition as in (3.2) can be shown to hold only if t_e is sufficiently small. Hence in that case existence and uniqueness of a solution is guaranteed only for a sufficiently short time interval.

The θ-scheme (3.3) can be applied to the ODE system (3.16), which results in

$$\frac{\vec{\mathbf{u}}^{n+1} - \vec{\mathbf{u}}^n}{\Delta t} = \theta\, \mathbf{P}\mathbf{M}^{-1}\, g(\vec{\mathbf{u}}^{n+1}, t_{n+1}) + (1-\theta)\mathbf{P}\mathbf{M}^{-1}\, g(\vec{\mathbf{u}}^n, t_n). \tag{3.18}$$

We assume that, for a given $\vec{\mathbf{u}}^0$, this recursion has a unique solution (which holds for Δt sufficiently small). In addition we assume that $\mathbf{B}\vec{\mathbf{u}}^0 = 0$ holds. From (3.18) and $\mathbf{B}\mathbf{P} = 0$ it then follows that $\mathbf{B}\vec{\mathbf{u}}^n = 0$ holds for all n. Based on (3.17) we introduce a pressure variable $\vec{\mathbf{p}}^k$ such that the corresponding finite element pressure function p_h satisfies $(p_h, 1)_{L^2} = 0$ and such that $\mathbf{B}\mathbf{M}^{-1}\mathbf{B}^T\vec{\mathbf{p}}^k = \mathbf{B}\mathbf{M}^{-1}\, g(\vec{\mathbf{u}}^k, t_k)$ holds. Using the definition of the projection $\mathbf{P}$ the recurrence relation in (3.18) can be rewritten as

$$\begin{aligned}\frac{\vec{\mathbf{u}}^{n+1} - \vec{\mathbf{u}}^n}{\Delta t} &+ \mathbf{M}^{-1}\mathbf{B}^T\big(\theta\vec{\mathbf{p}}^{n+1} + (1-\theta)\vec{\mathbf{p}}^n\big) \\ &= \theta\mathbf{M}^{-1}\, g(\vec{\mathbf{u}}^{n+1}, t_{n+1}) + (1-\theta)\mathbf{M}^{-1}\, g(\vec{\mathbf{u}}^n, t_n).\end{aligned}$$

Thus for given $\vec{\mathbf{u}}^n$ the pair $\vec{\mathbf{u}}^{n+1}$, $\vec{\tilde{\mathbf{p}}} := \theta\vec{\mathbf{p}}^{n+1} + (1-\theta)\vec{\mathbf{p}}^n$ is a solution of

$$\frac{\vec{\mathbf{u}}^{n+1} - \vec{\mathbf{u}}^n}{\Delta t} + \mathbf{M}^{-1}\mathbf{B}^T\vec{\tilde{\mathbf{p}}} = \theta\mathbf{M}^{-1}\, g(\vec{\mathbf{u}}^{n+1}, t_{n+1}) + (1-\theta)\mathbf{M}^{-1}\, g(\vec{\mathbf{u}}^n, t_n), \tag{3.19}$$

$$\mathbf{B}\,\vec{\mathbf{u}}^{n+1} = 0, \tag{3.20}$$

For given $\vec{\mathbf{u}}^n$ this saddle point problem has (for Δt sufficiently small) a unique solution pair $(\vec{\mathbf{u}}^{n+1}, \vec{\tilde{\mathbf{p}}})$ (on the subspace of pressure functions that satisfy $(p_h, 1)_{L^2} = 0$). Thus instead of (3.18) for computing $\vec{\mathbf{u}}^{n+1}$ we can use the equivalent formulation in (3.19)-(3.20) for computing $\vec{\mathbf{u}}^{n+1}$, $\vec{\tilde{\mathbf{p}}}$. An important advantage of the latter formulation is that the projection $\mathbf{P}$ has been eliminated. In the derivation it is essential that the mass matrix $\mathbf{M}$ does not depend on t. Summarizing, the θ-method for the Navier-Stokes DAE system takes the form

$$\begin{aligned} \mathbf{M}\frac{\vec{\mathbf{u}}^{n+1}-\vec{\mathbf{u}}^{n}}{\Delta t} + \theta[\mathbf{A}\vec{\mathbf{u}}^{n+1} + \mathbf{N}(\vec{\mathbf{u}}^{n+1})\vec{\mathbf{u}}^{n+1}] + \mathbf{B}^T\vec{\tilde{\mathbf{p}}} \\ = \theta\vec{\mathbf{f}}^{n+1} - (1-\theta)[\mathbf{A}\vec{\mathbf{u}}^{n} + \mathbf{N}(\vec{\mathbf{u}}^{n})\vec{\mathbf{u}}^{n} - \vec{\mathbf{f}}^{n}] \\ \mathbf{B}\vec{\mathbf{u}}^{n+1} = 0. \end{aligned} \tag{3.21}$$

The θ-schema applied to the Stokes problem results in a system as in (3.21), with the two terms $\mathbf{N}(\cdot)$ replaced by 0. In each time step a system of equations for the unknowns $\vec{\mathbf{u}}^{n+1}, \vec{\tilde{\mathbf{p}}}$ has to be solved. For the (Navier-)Stokes problem this saddle point system is (non)linear.

Remark 3.2.2 In the derivation above, we applied the *method of lines* approach, in which we first discretize the space variable and then the time variable. We comment on an alternative approach, often called *Rothe's method*, in which first the time variable and then the space variable is discretized. We explain this for the Stokes case. Starting point is the time dependent Stokes problem in which the pressure has been eliminated, i.e. a formulation as in (1.46). This is a variational formulation of an ODE in the function space $\mathbf{V}_{\mathrm{div}}$. To this problem one can apply the θ-scheme for discretization of the time variable, resulting in the following problem: given $\mathbf{u}^0 \in \mathbf{V}_{\mathrm{div}}$, for $n \geq 0$, determine $\mathbf{u}^{n+1} \in \mathbf{V}_{\mathrm{div}}$ such that

$$\begin{aligned} &\frac{1}{\Delta t}(\mathbf{u}^{n+1}-\mathbf{u}^{n},\mathbf{v})_{L^2} + \theta a(\mathbf{u}^{n+1},\mathbf{v}) \\ &= \theta\mathbf{g}^{n+1} - (1-\theta)\big[a(\mathbf{u}^{n},\mathbf{v}) - \mathbf{g}^{n}\big] \quad \text{for all} \quad \mathbf{v} \in \mathbf{V}_{\mathrm{div}}. \end{aligned} \tag{3.22}$$

This is a "projected" (due to $\mathbf{V}_{\mathrm{div}}$) *stationary* Stokes problem for the unknown function $\mathbf{u}^{n+1}$. Since finite element subspaces of $\mathbf{V}_{\mathrm{div}}$ are in general difficult to construct, we reformulate this problem as a saddle point problem in $\mathbf{V} \times Q = H_0^1(\Omega)^3 \times L_0^2(\Omega)$. Define the bilinear form

$$\hat{a}(\mathbf{u},\mathbf{v}) = \frac{1}{\Delta t}(\mathbf{u},\mathbf{v})_{L^2} + \theta a(\mathbf{u},\mathbf{v}), \quad \mathbf{u},\mathbf{v} \in \mathbf{V},\ \theta \in (0,1].$$

This bilinear form is elliptic and continuous on $\mathbf{V}$. We can apply the abstract theory in Section 4.3, Theorems 4.3.1 and 4.3.4, from which it follows that the problem (3.22) has a unique solution $\mathbf{u}^{n+1}$ which can also be characterized by the following Oseen problem: determine $\mathbf{u}^{n+1} \in \mathbf{V}$ and $p \in Q$ such that

$$\begin{aligned} &\frac{1}{\Delta t}(\mathbf{u}^{n+1}-\mathbf{u}^{n},\mathbf{v})_{L^2} + \theta a(\mathbf{u}^{n+1},\mathbf{v}) + b(\mathbf{v},p) \\ &= \theta\mathbf{g}^{n+1} - (1-\theta)\big[a(\mathbf{u}^{n},\mathbf{v}) - \mathbf{g}^{n}\big] \quad \text{for all} \quad \mathbf{v} \in \mathbf{V}, \\ &b(\mathbf{u},q) = 0 \quad \text{for all} \quad q \in Q. \end{aligned} \tag{3.23}$$

To this problem we can apply a Galerkin discretization with spaces $\mathbf{V}_h \subset \mathbf{V}$, $Q_h \subset Q$. Using standard nodal bases, we then obtain a fully discrete problem as in (3.21), with $N(\cdot) = 0$. The mass and stiffness matrices $\mathbf{M}$ and $\mathbf{A}$ and the right-hand sides $\vec{\mathbf{f}}^n$ are the same in the two approaches. Hence, the two methods yield the same results.

Although these two approaches turn out to be equivalent in case of an non-stationary Stokes problem with Hood-Taylor finite element spaces for spatial discretization and the θ-scheme for time discretization we comment on a subtle difference between the methods that will become important if we treat two-phase flow problems. For the method of lines the approach is as follows: we start with a saddle point problem for $(\mathbf{u},p)$, apply spatial Galerkin discretization, eliminate p_h, apply time discretization, introduce p_h again. For Rothe's method: start with a saddle point problem for $(\mathbf{u},p)$, eliminate p, apply time discretization, introduce p again, apply spatial Galerkin discretization. We see that in the former method we eliminate and re-introduce the spatially discrete pressure variable p_h, whereas in the latter this is done for the spatially continuous variable p. If in a time step $t_n \to t_{n+1}$ one wants to use *different* pressure

finite element spaces, then this pressure elimination and re-introduction can be problematic for the method of lines approach, whereas this is not the case for Rothe's method.

3.3 Fractional-step θ-scheme for the Navier-Stokes problem

Applying the fractional-step θ-scheme to the Navier-Stokes problem in ODE form (3.16) and transforming it back to its original DAE form along the same lines as in Section 3.2 results in

$$\left\{ \begin{array}{l} \mathbf{M}\frac{\vec{\mathbf{u}}^{n+\theta}-\vec{\mathbf{u}}^{n}}{\theta\Delta t} + \alpha[\mathbf{A}\vec{\mathbf{u}}^{n+\theta} + \mathbf{N}(\vec{\mathbf{u}}^{n+\theta})\vec{\mathbf{u}}^{n+\theta}] + \mathbf{B}^T\vec{\tilde{\mathbf{p}}}^{1} \\ \qquad = \vec{\mathbf{f}}^{n} - (1-\alpha)[\mathbf{A}\vec{\mathbf{u}}^{n} + \mathbf{N}(\vec{\mathbf{u}}^{n})\vec{\mathbf{u}}^{n}] \\ \mathbf{B}\vec{\mathbf{u}}^{n+\theta} = 0 \end{array} \right. \tag{3.24}$$

$$\left\{ \begin{array}{l} \mathbf{M}\frac{\vec{\mathbf{u}}^{n+1-\theta}-\vec{\mathbf{u}}^{n+\theta}}{(1-2\theta)\Delta t} + (1-\alpha)[\mathbf{A}\vec{\mathbf{u}}^{n+1-\theta} + \mathbf{N}(\vec{\mathbf{u}}^{n+1-\theta})\vec{\mathbf{u}}^{n+1-\theta}] + \mathbf{B}^T\vec{\tilde{\mathbf{p}}}^{2} \\ \qquad = \vec{\mathbf{f}}^{n+1-\theta} - \alpha[\mathbf{A}\vec{\mathbf{u}}^{n+\theta} + \mathbf{N}(\vec{\mathbf{u}}^{n+\theta})\vec{\mathbf{u}}^{n+\theta}] \\ \mathbf{B}\vec{\mathbf{u}}^{n+1-\theta} = 0 \end{array} \right. \tag{3.25}$$

$$\left\{ \begin{array}{l} \mathbf{M}\frac{\vec{\mathbf{u}}^{n+1}-\vec{\mathbf{u}}^{n+1-\theta}}{\theta\Delta t} + \alpha[\mathbf{A}\vec{\mathbf{u}}^{n+1} + \mathbf{N}(\vec{\mathbf{u}}^{n+1})\vec{\mathbf{u}}^{n+1}] + \mathbf{B}^T\vec{\tilde{\mathbf{p}}}^{3} \\ \qquad = \vec{\mathbf{f}}^{n+1-\theta} - (1-\alpha)[\mathbf{A}\vec{\mathbf{u}}^{n+1-\theta} + \mathbf{N}(\vec{\mathbf{u}}^{n+1-\theta})\vec{\mathbf{u}}^{n+1-\theta}] \\ \mathbf{B}\vec{\mathbf{u}}^{n+1} = 0 \end{array} \right. \tag{3.26}$$

If we take parameter values as in (3.14) then the *non*linear problems for the pairs $(\vec{\mathbf{u}}^{n+\theta}, \vec{\tilde{\mathbf{p}}}^{1})$, $(\vec{\mathbf{u}}^{n+1-\theta}, \vec{\tilde{\mathbf{p}}}^{2})$, $(\vec{\mathbf{u}}^{n+1}, \vec{\tilde{\mathbf{p}}}^{3})$ in these three substeps have a similar form. We obtain the fractional-step θ-scheme for the Stokes by replacing all terms $\mathbf{N}(\cdot)$ by 0.

Remark 3.3.1 If one uses the variant of the fractional-step θ-scheme as described in Remark 3.1.3 then in each time interval $[n\Delta t, (n+1)\Delta t]$ three successive substeps of the θ-scheme (3.21) (with different values for θ) are applied.

3.4 Numerical experiments

To analyze the time discretization error for different time integration schemes, we reconsider the test case of a rectangular tube described in Section 2.2.1. Instead of a stationary Stokes problem we now consider the non-stationary Stokes problem (1.47) on $\Omega \times [0,T]$ with $T = 2$ for different time step sizes Δt. To obtain a time-dependent velocity and pressure field, we prescribe an oscillating boundary condition $\mathbf{u}(0, x_2, x_3) = s(x_2, x_3)\big(1 + 0.25\sin(2\pi t)\big)$ at the inflow boundary $x_1 = 0$, with s defined in (2.38), an outflow boundary condition $\boldsymbol{\sigma}\mathbf{n} = 0$ for $x_1 = L$ and $\mathbf{u} = 0$ on the remaining boundaries.

For spatial discretization we use the Hood-Taylor finite element pair for $k = 2$ on a triangulation $\mathcal{T}$ which is constructed by subdividing Ω into $16 \times 4 \times 4$ sub-cubes each consisting of 6 tetrahedra. For this fixed spatial discretization different time integration schemes are analyzed for different time step sizes $\Delta t = T/n_t$ where $n_t = 25, 50, 100, 200, 400, 800$ denotes the number of time steps applied to obtain the approximations $\vec{\mathbf{u}}^{n_t}$, $\vec{p}^{\,n_t}$ to $\vec{\mathbf{u}}(T)$, $\vec{p}(T)$, respectively. As the exact solutions $\vec{\mathbf{u}}(T), \vec{p}(T)$ of the DAE system (2.21) are not available we instead use reference solutions $\vec{\mathbf{u}}^{\mathrm{ref}}$, $\vec{p}^{\,\mathrm{ref}}$ obtained by applying 2000 steps of the fractional-step θ-scheme with step size $\Delta t = 10^{-3}$.

For a fixed spatial coordinate $x = (2, 0.5, 0.5)$ in the center of the domain Ω, the first velocity component $u_1(x,t)$ and pressure $p(x,t)$ are shown as a function of time $t \in [0,2]$ in Figure 3.3. Also given are the results for the implicit Euler scheme ($\theta = 1$), the Crank-Nicolson scheme ($\theta = 0.5$) and the fractional-step θ-scheme applying 25 steps with time step

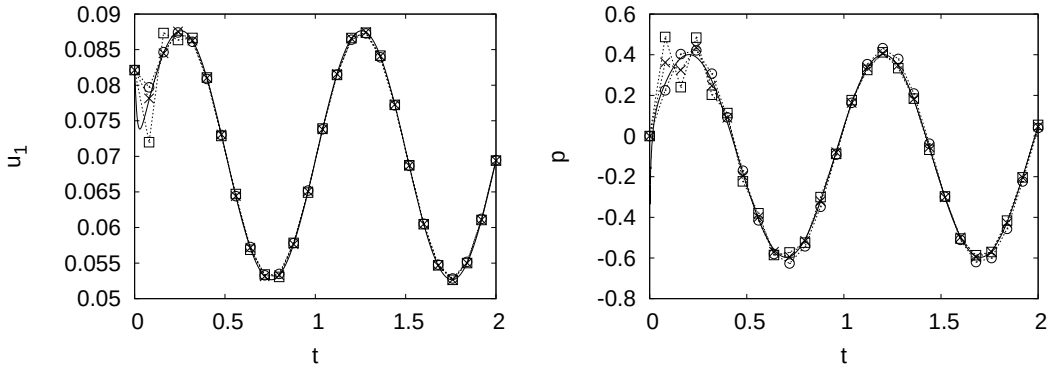

Fig. 3.3. Velocity u_1 (left) and pressure p (right) at point $x = (2, 0.5, 0.5) \in \Omega$ as a function of time. Shown are the reference solution (solid line) and implicit Euler (circles), Crank-Nicolson (squares) and fractional-step (crosses) solutions for 25 time steps, respectively.

n_t	$\|\vec{\mathbf{u}}^{\mathrm{ref}} - \vec{\mathbf{u}}^{n_t}\|_{L^2}$	order	$\|\vec{\mathbf{u}}^{\mathrm{ref}} - \vec{\mathbf{u}}^{n_t}\|_1$	order	$\|\vec{p}^{\,\mathrm{ref}} - \vec{p}^{\,n_t}\|_{L^2}$	order
25	3.69 E-5	—	3.54 E-4	—	7.71 E-3	—
50	1.38 E-5	1.42	1.32 E-4	1.43	2.17 E-3	1.83
100	5.69 E-6	1.29	5.40 E-5	1.29	6.51 E-4	1.73
200	2.53 E-6	1.17	2.40 E-5	1.17	2.17 E-4	1.59
400	1.19 E-6	1.09	1.12 E-5	1.09	8.13 E-5	1.42
800	5.75 E-7	1.05	5.43 E-6	1.05	3.38 E-5	1.26

Table 3.1. Convergence behavior of the implicit Euler scheme w.r.t. time step size.

n_t	$\|\vec{\mathbf{u}}^{\mathrm{ref}} - \vec{\mathbf{u}}^{n_t}\|_{L^2}$	order	$\|\vec{\mathbf{u}}^{\mathrm{ref}} - \vec{\mathbf{u}}^{n_t}\|_1$	order	$\|\vec{p}^{\,\mathrm{ref}} - \vec{p}^{\,n_t}\|_{L^2}$	order
25	2.75 E-5	—	6.87 E-4	—	9.16 E-3	—
50	3.84 E-6	2.84	1.38 E-4	2.31	9.65 E-5	6.57
100	6.56 E-7	2.55	6.66 E-6	4.38	4.21 E-5	1.20
200	1.63 E-7	2.00	1.60 E-6	2.06	1.04 E-5	2.02
400	4.07 E-8	2.01	3.98 E-7	2.01	2.57 E-6	2.01
800	9.99 E-9	2.03	9.78 E-8	2.03	6.56 E-7	1.97

Table 3.2. Convergence behavior of the Crank-Nicolson scheme w.r.t. time step size.

n_t	$\|\vec{\mathbf{u}}^{\mathrm{ref}} - \vec{\mathbf{u}}^{n_t}\|_{L^2}$	order	$\|\vec{\mathbf{u}}^{\mathrm{ref}} - \vec{\mathbf{u}}^{n_t}\|_1$	order	$\|\vec{p}^{\,\mathrm{ref}} - \vec{p}^{\,n_t}\|_{L^2}$	order
25	5.85 E-7	—	7.93 E-6	—	2.67 E-4	—
50	3.40 E-7	0.78	3.31 E-6	1.26	5.50 E-5	2.28
100	9.49 E-8	1.84	9.08 E-7	1.87	1.14 E-5	2.27
200	2.37 E-8	2.00	2.30 E-7	1.98	2.56 E-6	2.15
400	5.70 E-9	2.06	5.58 E-8	2.05	6.02 E-7	2.09
800	1.24 E-9	2.20	1.21 E-9	2.20	1.31 E-7	2.20

Table 3.3. Convergence behavior of the fractional-step θ-scheme w.r.t. time step size.

size $\Delta t = 0.08$. We notice an oscillatory behavior of the Crank-Nicolson scheme in the first time steps which is probably due to the fact that this method does not have a good smoothing property, as explained in Section 3.1.

Tables 3.1–3.3 show the convergence w.r.t. time step size for the different time discretization schemes. The numerical experiments confirm the first order convergence of the implicit Euler scheme and second order convergence of the Crank-Nicolson and the fractional-step θ-scheme.

Comparing the second order schemes we observe that the errors for the fractional-step scheme are smaller than those of the Crank-Nicolson scheme by about a factor of 10. Note, however, that in the fractional-step scheme three macro-steps are performed per time step, and thus for a fair comparison it should be compared to a Crank-Nicolson scheme with a time step size divided by 3. This would lead to Crank-Nicolson errors which are roughly $3^2 = 9$ times smaller than those in Table 3.2 and thus are of the same order of magnitude as the errors in the fractional-step scheme given in Table 3.3.

3.5 Discussion and additional references

In this chapter we restricted ourselves to basic, but still very popular, time discretization methods for non-stationary (Navier-)Stokes equations. We briefly discuss a few related aspects.

Error analyses of a fully (space and time) discrete problem as in (3.21) are presented in [25, 26, 27]. In the literature there are only very few studies on *adaptive* time stepping for solving non-stationary Navier-Stokes equations; a recent paper is [29]. There are several variants of the fractional-step θ-scheme for the Navier-Stokes equations based on different operator splittings. Some of these are discussed in [21]. A popular variant is based on a semi-implict treatment of the nonlinear term in the Navier-Stokes equations. In such a method one replaces the term $\mathbf{N}(\vec{\mathbf{u}}^{n+\theta})\vec{\mathbf{u}}^{n+\theta}$ in (3.24) by $\mathbf{N}(\vec{\mathbf{u}}^{n})\vec{\mathbf{u}}^{n+\theta}$, and similarly for the nonlinear terms in (3.25), (3.26), cf. [34]. Alternatively, instead of replacing $\vec{\mathbf{u}}^{n+\theta}$ by $\vec{\mathbf{u}}^{n}$ one can also replace it by a more accurate extrapolation of $\vec{\mathbf{u}}^{n}$ and $\vec{\mathbf{u}}^{n-1}$. Other semi-implicit methods are explained in [34].

A class of methods that is particularly popular in the engineering literature are the so-called projection methods. These methods have a predictor-corrector structure, in which in the predictor step, which does not involve pressure, a new velocity field is determined and in the corrector step, which involves a pressure variable, this new velocity field is "projected" onto the subspace of divergence free functions. To explain the main idea we consider a basic variant of this method in semi-discrete form only, i.e. we discretize in time but not in space, and formulate it in strong formulation. Let $\mathbf{u}^n \in \mathbf{V} := H_0^1(\Omega)^3$ be a given approximation of $\mathbf{u}(\cdot, t_n)$. We define $\tilde{\mathbf{u}}^{n+1} \in \mathbf{V}$ as the solution of

$$\frac{1}{\Delta t}\big(\tilde{\mathbf{u}}^{n+1} - \mathbf{u}^n\big) - \frac{1}{Re}\Delta\tilde{\mathbf{u}}^{n+1} + (\mathbf{u}^n \cdot \nabla)\tilde{\mathbf{u}}^{n+1} = \mathbf{g}(t_{n+1}).$$

The approximation $\tilde{\mathbf{u}}^{n+1} \approx \mathbf{u}(\cdot, t_{n+1})$ is projected onto the space of divergence free functions by solving a saddle point problem: determine $\mathbf{u}^{n+1} \in \mathbf{V}$ and $q \in L_0^2(\Omega)$ such that

$$\begin{aligned} \frac{1}{\Delta t}\big(\mathbf{u}^{n+1} - \tilde{\mathbf{u}}^{n+1}\big) + \nabla q = 0 \quad &\text{in } \Omega \\ \operatorname{div}\mathbf{u}^{n+1} = 0 \quad &\text{in } \Omega \\ \mathbf{u}^{n+1}\cdot\mathbf{n} = 0 \quad &\text{on } \partial\Omega. \end{aligned}$$

An detailed study of this projection method and variants of it can be found in [32].

4

Appendix: Variational formulations in Hilbert spaces

In this appendix we collect some results on the well-posedness of variational problems in Hilbert spaces. These results are known in the literature. For most of the proofs we refer to the literature.

4.1 Variational problems and Galerkin discretization

We start with a remark on notation: In this appendix, for elements from a Hilbert space we use boldface notation (e.g., $\mathbf{u}$), elements from the dual space (i.e., bounded linear functionals) are denoted by f, g, etc., and for linear operators between spaces we use capitals (e.g., L).

Let H_1 and H_2 be Hilbert spaces. A bilinear form $k : H_1 \times H_2 \to \mathbb{R}$ is *continuous* if there is a constant M such that for all $\mathbf{x} \in H_1,\ \mathbf{y} \in H_2$:

$$|k(\mathbf{x}, \mathbf{y})| \leq M \|\mathbf{x}\|_{H_1} \|\mathbf{y}\|_{H_2}. \tag{4.1}$$

For a continuous bilinear form $k : H_1 \times H_2 \to \mathbb{R}$ we define its norm by $\|k\| = \sup \{\, |k(\mathbf{x}, \mathbf{y})| : \|\mathbf{x}\|_{H_1} = 1,\ \|\mathbf{y}\|_{H_2} =$
A fundamental result is given in the following theorem:

Theorem 4.1.1 *Let H_1, H_2 be Hilbert spaces and $k : H_1 \times H_2 \to \mathbb{R}$ be a continuous bilinear form. For $f \in H_2'$ consider the variational problem:*

$$\textit{find} \quad \mathbf{u} \in H_1 \quad \textit{such that} \quad k(\mathbf{u}, \mathbf{v}) = f(\mathbf{v}) \quad \textit{for all} \quad \mathbf{v} \in H_2. \tag{4.2}$$

The following two statements are equivalent*:*

1. *For arbitrary $f \in H_2'$ the problem* (4.2) *has a unique solution $\mathbf{u} \in H_1$ and $\|\mathbf{u}\|_{H_1} \leq c\|f\|_{H_2'}$ holds with a constant c independent of f.*
2. *The conditions* (4.3) *and* (4.4) *hold:*

$$\exists\ \varepsilon > 0\ :\quad \sup_{\mathbf{v} \in H_2} \frac{k(\mathbf{u}, \mathbf{v})}{\|\mathbf{v}\|_{H_2}} \geq \varepsilon \|\mathbf{u}\|_{H_1} \quad \textit{for all} \quad \mathbf{u} \in H_1, \tag{4.3}$$

$$\forall\ \mathbf{v} \in H_2,\ \mathbf{v} \neq 0,\quad \exists\ \ \mathbf{u} \in H_1\ :\quad k(\mathbf{u}, \mathbf{v}) \neq 0. \tag{4.4}$$

Moreover, for the constants c and ε one can take $c = \frac{1}{\varepsilon}$.

A proof of this result can be found in e.g. [16].

Remark 4.1.2 The condition (4.4) can also be formulated as follows:

$$\big[\mathbf{v} \in H_2 \ \text{ such that } \ k(\mathbf{u}, \mathbf{v}) = 0 \ \text{ for all } \ \mathbf{u} \in H_1\big] \ \Rightarrow\ \mathbf{v} = 0.$$

The condition (4.3) is equivalent to

$$\exists\, \varepsilon > 0 \;:\quad \inf_{\mathbf{u}\in H_1\setminus\{0\}} \sup_{\mathbf{v}\in H_2} \frac{k(\mathbf{u},\mathbf{v})}{\|\mathbf{u}\|_{H_1}\|\mathbf{v}\|_{H_2}} \geq \varepsilon, \tag{4.5}$$

and is often called the *inf-sup condition*. In the *finite* dimensional case with $\dim(H_1) = \dim(H_2) < \infty$ this condition implies the result in (4.4) and thus is *necessary and sufficient* for existence and uniqueness.

The Galerkin discretization of the problem (4.2) is based on the following simple idea. We assume *finite dimensional subspaces* $H_{1,h} \subset H_1$, $H_{2,h} \subset H_2$ (note: in concrete cases the index h will correspond to some mesh size parameter) and consider the finite dimensional variational problem

$$\text{find } \mathbf{u}_h \in H_{1,h} \text{ such that } k(\mathbf{u}_h,\mathbf{v}_h) = f(\mathbf{v}_h) \text{ for all } \mathbf{v}_h \in H_{2,h}. \tag{4.6}$$

This problem is called a *Galerkin discretization* of (4.2) (in $H_{1,h} \times H_{2,h}$). We now discuss the well-posedness of this Galerkin-discretization. First note that the continuity of $k : H_{1,h} \times H_{2,h} \to \mathbb{R}$ follows from (4.1). From theorem 4.1.1 it follows that we need the conditions (4.3) and (4.4) with H_i replaced by $H_{i,h}$, $i = 1, 2$. However, because $H_{i,h}$ is finite dimensional we only need (4.3) since this implies (4.4). Thus we formulate the following (discrete) inf-sup condition in the space $H_{1,h} \times H_{2,h}$:

$$\exists\, \varepsilon_h > 0 \;:\quad \sup_{\mathbf{v}_h\in H_{2,h}} \frac{k(\mathbf{u}_h,\mathbf{v}_h)}{\|\mathbf{v}_h\|_{H_2}} \geq \varepsilon_h \,\|\mathbf{u}_h\|_{H_1} \quad \text{for all} \quad \mathbf{u}_h \in H_{1,h}. \tag{4.7}$$

We prove a fundamental result in which the *discretization* error $\|\mathbf{u} - \mathbf{u}_h\|_{H_1}$ is bounded by an *approximation* error $\inf_{\mathbf{v}_h\in H_{1,h}} \|\mathbf{u} - \mathbf{v}_h\|_{H_1}$. In the literature this result is often called "Céa's lemma".

Theorem 4.1.3 (Céa's lemma.) *Let H_1, H_2 be Hilbert spaces and $k : H_1 \times H_2 \to \mathbb{R}$ be a bilinear form. Assume that* (4.1), (4.3), (4.4), (4.7) *hold. Then the variational problem* (4.2) *and its Galerkin discretization* (4.6) *have unique solutions* $\mathbf{u}$ *and* $\mathbf{u}_h$*, respectively. Furthermore, the inequality*

$$\|\mathbf{u} - \mathbf{u}_h\|_{H_1} \leq \big(1 + \frac{M}{\varepsilon_h}\big) \inf_{\mathbf{v}_h\in H_{1,h}} \|\mathbf{u} - \mathbf{v}_h\|_{H_1} \tag{4.8}$$

holds.

Proof. The existence and uniqueness of $\mathbf{u}$ and $\mathbf{u}_h$ follow from Theorem 4.1.1 and the fact that in the finite dimensional case (4.3) implies (4.4). From (4.2) and (4.6) it follows that

$$k(\mathbf{u} - \mathbf{u}_h, \mathbf{v}_h) = 0 \quad \text{for all} \quad \mathbf{v}_h \in H_{2,h}. \tag{4.9}$$

For arbitrary $\mathbf{v}_h \in H_{1,h}$ we have, due to (4.7), (4.9), (4.1):

$$\begin{aligned}\|\mathbf{v}_h - \mathbf{u}_h\|_{H_1} &\leq \frac{1}{\varepsilon_h} \sup_{\mathbf{w}_h\in H_{2,h}} \frac{k(\mathbf{v}_h - \mathbf{u}_h, \mathbf{w}_h)}{\|\mathbf{w}_h\|_{H_2}} \\ &= \frac{1}{\varepsilon_h} \sup_{\mathbf{w}_h\in H_{2,h}} \frac{k(\mathbf{v}_h - \mathbf{u}, \mathbf{w}_h)}{\|\mathbf{w}_h\|_{H_2}} \leq \frac{M}{\varepsilon_h}\|\mathbf{v}_h - \mathbf{u}\|_{H_1}.\end{aligned}$$

From this and the triangle inequality

$$\|\mathbf{u} - \mathbf{u}_h\|_{H_1} \leq \|\mathbf{u} - \mathbf{v}_h\|_{H_1} + \|\mathbf{v}_h - \mathbf{u}_h\|_{H_1} \quad \text{for all} \quad \mathbf{v}_h \in H_{1,h}$$

the result follows. □

4.2 Application to elliptic problems

In this section we apply the results from section 4.1 in the special case $H_1 = H_2 =: H$ and with a bilinear form $k : H \times H \to \mathbb{R}$ that is assumed to be *H-elliptic*, i.e., there exists a constant $\gamma > 0$ such

$$k(\mathbf{u}, \mathbf{u}) \geq \gamma \|\mathbf{u}\|_H^2 \quad \text{for all} \quad \mathbf{u} \in H.$$

From this property it follows that the conditions (4.3), (4.4) and (4.7) are satisfied with $\varepsilon = \varepsilon_h = \gamma$. Thus, as an immediate consequence of Theorem 4.1.1 we obtain the following famous result, often called "Lax-Milgram lemma".

Theorem 4.2.1 (Lax-Milgram lemma) *Let H be a Hilbert space and $k : H \times H \to \mathbb{R}$ a continuous H-elliptic bilinear form with ellipticity constant γ. Then for every $f \in H'$ there exists a unique $\mathbf{u} \in H$ such that*

$$k(\mathbf{u}, \mathbf{v}) = f(\mathbf{v}) \quad \textit{for all} \quad \mathbf{v} \in H. \tag{4.10}$$

Furthermore, the inequality $\|\mathbf{u}\|_H \leq \frac{1}{\gamma}\|f\|_{H'}$ holds.

If the bilinear form is in addition *symmetric*, i.e., $k(\mathbf{u}, \mathbf{v}) = k(\mathbf{v}, \mathbf{u})$ for all $\mathbf{u}, \mathbf{v} \in H$, then there is a natural correspondence between the variational problem (4.10) and a minimization problem:

Theorem 4.2.2 *Let H be a Hilbert space and $k : H \times H \to \mathbb{R}$ a continuous H-elliptic* symmetric *bilinear form. For $f \in H'$ let $\mathbf{u} \in H$ be the unique solution of the variational problem* (4.10)*. Then $\mathbf{u}$ is the unique minimizer of the functional*

$$J(\mathbf{v}) := \frac{1}{2}k(\mathbf{v}, \mathbf{v}) - f(\mathbf{v}). \tag{4.11}$$

Proof. From the Lax-Milgram lemma it follows that the variational problem (4.10) has a unique solution $\mathbf{u} \in H$. For arbitrary $\mathbf{z} \in H$, $\mathbf{z} \neq 0$, we have, with ellipticity constant $\gamma > 0$:

$$\begin{aligned} J(\mathbf{u} + \mathbf{z}) &= \frac{1}{2}k(\mathbf{u} + \mathbf{z}, \mathbf{u} + \mathbf{z}) - f(\mathbf{u} + \mathbf{z}) \\ &= \frac{1}{2}k(\mathbf{u}, \mathbf{u}) - f(\mathbf{u}) + k(\mathbf{u}, \mathbf{z}) - f(\mathbf{z}) + \frac{1}{2}k(\mathbf{z}, \mathbf{z}) \\ &= J(\mathbf{u}) + \frac{1}{2}k(\mathbf{z}, \mathbf{z}) \geq J(\mathbf{u}) + \frac{1}{2}\gamma\|\mathbf{z}\|_H^2 > J(\mathbf{u}). \end{aligned}$$

This proves the desired result. □

In the elliptic case one can improve the discretization error bound in Céa's lemma. First we give a result in which the term $1 + \frac{M}{\varepsilon_h} = 1 + \frac{M}{\gamma}$ is replaced by $\frac{M}{\gamma}$.

Theorem 4.2.3 *Consider the problem* (4.10) *and its Galerkin discretization in the subspace $H_h \subset H$. Assume that the conditions as in Theorem 4.2.1 are satisfied. Then the variational problem* (4.10) *and its Galerkin discretization have unique solutions $\mathbf{u}$ and $\mathbf{u}_h$, respectively. Furthermore, the inequality*

$$\|\mathbf{u} - \mathbf{u}_h\|_H \leq \frac{M}{\gamma} \inf_{\mathbf{v}_h \in H_h} \|\mathbf{u} - \mathbf{v}_h\|_H \tag{4.12}$$

holds.

Proof. Theorem 4.2.1 can be applied both to (4.10) and its Galerkin discretization. Thus we conclude that unique solutions $\mathbf{u}$ of (4.10) and $\mathbf{u}_h$ of the Galerkin discretization exist. Using

$k(\mathbf{u}-\mathbf{u}_h,\mathbf{v}_h)=0$ for all $\mathbf{v}_h\in H_h$ and the ellipticity and continuity properties, we get for arbitrary $\mathbf{v}_h\in H_h$:

$$\begin{aligned}\|\mathbf{u}-\mathbf{u}_h\|_H^2 &\le \frac{1}{\gamma}k(\mathbf{u}-\mathbf{u}_h,\mathbf{u}-\mathbf{u}_h)=\frac{1}{\gamma}k(\mathbf{u}-\mathbf{u}_h,\mathbf{u}-\mathbf{v}_h)\\ &\le \frac{M}{\gamma}\|\mathbf{u}-\mathbf{u}_h\|_H\|\mathbf{u}-\mathbf{v}_h\|_H.\end{aligned}$$

Hence the inequality in (4.12) holds. □

An improvement of the bound in (4.12) can be obtained if $k(\cdot,\cdot)$ is symmetric:

Theorem 4.2.4 *Assume that the conditions as in Theorem 4.2.3 are satisfied. If in addition the bilinear form $k(\cdot,\cdot)$ is* symmetric*, the inequality*

$$\|\mathbf{u}-\mathbf{u}_h\|_H \le \sqrt{\frac{M}{\gamma}}\inf_{\mathbf{v}_h\in H_h}\|\mathbf{u}-\mathbf{v}_h\|_H \tag{4.13}$$

holds.

Proof. Introduce the norm $|\!|\!|v|\!|\!| := k(\mathbf{v},\mathbf{v})^{\frac{1}{2}}$ on H. Note that

$$\sqrt{\gamma}\|\mathbf{v}\|_H \le |\!|\!|v|\!|\!| \le \sqrt{M}\|\mathbf{v}\|_H \quad \text{for all } \ \mathbf{v}\in H.$$

The space $(H,|\!|\!|\cdot|\!|\!|)$ is a Hilbert space and due to $|\!|\!|\mathbf{v}|\!|\!|^2 = k(\mathbf{v},\mathbf{v})$, $k(\mathbf{u},\mathbf{v}) \le |\!|\!|\mathbf{u}|\!|\!||\!|\!|\mathbf{v}|\!|\!|$ the bilinear form has ellipticity constant and continuity constant w.r.t. the norm $|\!|\!|\cdot|\!|\!|$ both equal to 1. Application of Theorem 4.2.3 in the space $(H,|\!|\!|\cdot|\!|\!|)$ yields

$$|\!|\!|\mathbf{u}-\mathbf{u}_h|\!|\!| \le \inf_{\mathbf{v}_h\in H_h}|\!|\!|\mathbf{u}-\mathbf{v}_h|\!|\!|$$

and thus we obtain

$$\begin{aligned}\|\mathbf{u}-\mathbf{u}_h\|_H &\le \frac{1}{\sqrt{\gamma}}|\!|\!|\mathbf{u}-\mathbf{u}_h|\!|\!| \le \frac{1}{\sqrt{\gamma}}\inf_{\mathbf{v}_h\in H_h}|\!|\!|\mathbf{u}-\mathbf{v}_h|\!|\!|\\ &\le \sqrt{\frac{M}{\gamma}}\inf_{\mathbf{v}_h\in H_h}\|\mathbf{u}-\mathbf{v}_h\|_H,\end{aligned}$$

which completes the proof. □

4.3 Application to saddle point problems

We introduce an abstract saddle point problem. Let V and M be Hilbert spaces and

$$\hat{a}: V\times V\to\mathbb{R}, \quad \hat{b}: V\times M\to\mathbb{R},$$

be continuous bilinear forms. For $f_1\in V'$, $f_2\in M'$ we define the following variational problem: find $(\boldsymbol{\phi},\boldsymbol{\lambda})\in V\times M$ such that

$$\hat{a}(\boldsymbol{\phi},\boldsymbol{\psi})+\hat{b}(\boldsymbol{\psi},\boldsymbol{\lambda}) = f_1(\boldsymbol{\psi}) \quad \text{for all } \ \boldsymbol{\psi}\in V \tag{4.14a}$$

$$\hat{b}(\boldsymbol{\phi},\boldsymbol{\mu}) = f_2(\boldsymbol{\mu}) \quad \text{for all } \ \boldsymbol{\mu}\in M. \tag{4.14b}$$

This variational problem can be put in the general framework of section 4.1 as follows. Define $H := V\times M$ and

$$\begin{aligned} &k : H \times H \to \mathbb{R}, \qquad k(\mathbf{u},\mathbf{v}) := \hat{a}(\boldsymbol{\phi},\boldsymbol{\psi}) + \hat{b}(\boldsymbol{\phi},\boldsymbol{\mu}) + \hat{b}(\boldsymbol{\psi},\boldsymbol{\lambda}), \\ &\text{with } \ \mathbf{u} := (\boldsymbol{\phi},\boldsymbol{\lambda}), \ \ \mathbf{v} := (\boldsymbol{\psi},\boldsymbol{\mu}). \end{aligned} \tag{4.15}$$

On H we use the product norm $\|\mathbf{u}\|_H^2 = \|\boldsymbol{\phi}\|_V^2 + \|\boldsymbol{\lambda}\|_M^2$, for $\mathbf{u} = (\boldsymbol{\phi},\boldsymbol{\lambda}) \in H$. If we define $f \in H' = V' \times M'$ by $f(\boldsymbol{\phi},\boldsymbol{\lambda}) = f_1(\boldsymbol{\psi}) + f_2(\boldsymbol{\mu})$ then the problem (4.14) can be reformulated in the setting of theorem 4.1.1 as follows:

$$\text{find } \ \mathbf{u} \in H \ \text{ such that } \ k(\mathbf{u},\mathbf{v}) = f(\mathbf{v}) \quad \text{for all } \ \mathbf{v} \in H. \tag{4.16}$$

Based on Theorem 4.1.1 the following well-posedness result for the saddle point problem can be derived, cf. [8, 20], in which the conditions (4.3) and (4.4) on the bilinear form $k(\cdot,\cdot)$ are replaced by conditions on $\hat{a}(\cdot,\cdot)$ and $\hat{b}(\cdot,\cdot)$.

Theorem 4.3.1 *For arbitrary $f_1 \in V', f_2 \in M'$ consider the variational problem* (4.14). *Assume that the bilinear forms $\hat{a}(\cdot,\cdot)$ and $\hat{b}(\cdot,\cdot)$ are continuous and satisfy the following two conditions:*

$$\exists\ \beta > 0 : \ \sup_{\boldsymbol{\psi} \in V} \frac{\hat{b}(\boldsymbol{\psi},\boldsymbol{\lambda})}{\|\boldsymbol{\psi}\|_V} \geq \beta \, \|\boldsymbol{\lambda}\|_M \quad \forall\ \boldsymbol{\lambda} \in M \quad \text{(inf-sup)}, \tag{4.17a}$$

$$\exists\ \gamma > 0 \ : \quad \hat{a}(\boldsymbol{\phi},\boldsymbol{\phi}) \geq \gamma \, \|\boldsymbol{\phi}\|_V^2 \ \ \forall\ \boldsymbol{\phi} \in V \ \ \text{(V-ellipt.)}. \tag{4.17b}$$

Then the problem (4.14) *has a unique solution $(\boldsymbol{\phi},\boldsymbol{\lambda})$. Moreover, the stability bound*

$$\big(\|\boldsymbol{\phi}\|_V^2 + \|\boldsymbol{\lambda}\|_M^2\big)^{\frac{1}{2}} \leq \frac{(\beta + 2\|\hat{a}\|)^2}{\gamma\beta^2} \big(\|f_1\|_{V'}^2 + \|f_2\|_{M'}^2\big)^{\frac{1}{2}}$$

holds. Hence the problem (4.14) *is well-posed.*

Remark 4.3.2 The *inf-sup* condition in (4.17a) is not only sufficient but also *necessary* for well-posedness of the saddle point problem (4.14). The condition on $\hat{a}(\cdot,\cdot)$ in (4.17b) is sufficient but not necessary. It turns out that the following two conditions for $\hat{a}(\cdot,\cdot)$ together are necessary and sufficient:

$$\begin{aligned} \exists\ \delta > 0 : \quad \sup_{\boldsymbol{\psi} \in V_0} \frac{\hat{a}(\boldsymbol{\phi},\boldsymbol{\psi})}{\|\boldsymbol{\psi}\|_V} \geq \delta \, \|\boldsymbol{\phi}\|_V \quad &\text{for all } \ \boldsymbol{\phi} \in V_0, \\ \forall\ \boldsymbol{\psi} \in V_0, \ \boldsymbol{\psi} \neq 0, \ \ \exists\ \boldsymbol{\phi} \in V_0 \ : \ \ \hat{a}(\boldsymbol{\phi},\boldsymbol{\psi}) &\neq 0, \end{aligned}$$

with $V_0 := \Big\{ \boldsymbol{\phi} \in V : \ \hat{b}(\boldsymbol{\phi},\boldsymbol{\lambda}) = 0 \quad \text{for all } \ \boldsymbol{\lambda} \in M \Big\}$.

If $\hat{a}(\cdot,\cdot)$ is in addition assumed to be symmetric, then as in Theorem 4.2.2 there is a natural correspondence between the variational problem (4.14) and extrema of a functional:

Theorem 4.3.3 *Assume that the bilinear forms $\hat{a}(\cdot,\cdot)$ and $\hat{b}(\cdot,\cdot)$ are continuous and satisfy the conditions* (4.17). *In addition we assume that $\hat{a}(\cdot,\cdot)$ is* symmetric. *For arbitrary $f_1 \in V', f_2 \in M'$ let $(\boldsymbol{\phi},\boldsymbol{\lambda})$ be the unique solution of* (4.14). *Define the functional $\mathcal{L} : V \times M \to \mathbb{R}$ by*

$$\mathcal{L}(\boldsymbol{\psi},\boldsymbol{\mu}) = \frac{1}{2}\hat{a}(\boldsymbol{\psi},\boldsymbol{\psi}) + \hat{b}(\boldsymbol{\psi},\boldsymbol{\mu}) - f_1(\boldsymbol{\psi}) - f_2(\boldsymbol{\mu}).$$

Then $(\boldsymbol{\phi},\boldsymbol{\lambda})$ is also the unique element in $V \times M$ for which

$$\mathcal{L}(\boldsymbol{\phi},\boldsymbol{\mu}) \leq \mathcal{L}(\boldsymbol{\phi},\boldsymbol{\lambda}) \leq \mathcal{L}(\boldsymbol{\psi},\boldsymbol{\lambda}) \quad \textit{for all } \ \boldsymbol{\psi} \in V, \ \boldsymbol{\mu} \in M \tag{4.18}$$

holds.

For a proof of this result we refer to the literature, e.g. [20]. The property in (4.18) explains why this type of variational equations are called "saddle point" problems.

The unknown $\boldsymbol{\lambda}$ in (4.14) can be eliminated, resulting in an equivalent formulation in which only the unknown $\boldsymbol{\phi}$ occurs. For this we introduce the following notation, for $f_2 \in M'$:

$$V_{f_2} := \Big\{ \boldsymbol{\phi} \in V : \hat{b}(\boldsymbol{\phi}, \boldsymbol{\mu}) = f_2(\boldsymbol{\mu}) \quad \text{for all} \quad \boldsymbol{\mu} \in M \Big\}.$$

We consider the variational problem: determine $\boldsymbol{\phi} \in V_{f_2}$ such that

$$\hat{a}(\boldsymbol{\phi}, \boldsymbol{\psi}) = f_1(\boldsymbol{\psi}) \quad \text{for all} \quad \boldsymbol{\psi} \in V_0. \tag{4.19}$$

The equivalence of this problem and the saddle point problem (4.14) is given in the following theorem.

Theorem 4.3.4 *Let the assumptions as in Theorem 4.3.1 be satisfied. Let $(\boldsymbol{\phi}, \boldsymbol{\lambda})$ be the unique solution of problem* (4.14)*. Then $\boldsymbol{\phi}$ is the unique solution of the variational problem* (4.19)*.*

Proof. For $\boldsymbol{\phi}$ we have $\hat{b}(\boldsymbol{\phi}, \boldsymbol{\mu}) = f_2(\boldsymbol{\mu})$ for all $\boldsymbol{\mu} \in M$, hence $\boldsymbol{\phi} \in V_{f_2}$. From (4.14a) and $\hat{b}(\boldsymbol{\psi}, \boldsymbol{\lambda}) = 0$ for all $\boldsymbol{\psi} \in V_0$ it follows that

$$\hat{a}(\boldsymbol{\phi}, \boldsymbol{\psi}) = f_1(\boldsymbol{\psi}) \quad \text{for all} \quad \boldsymbol{\psi} \in V_0,$$

and thus $\boldsymbol{\psi}$ solves the problem (4.19). Uniqueness of this solution follows using the ellipticity property (4.17b). □

Now we consider the *Galerkin discretization* of the saddle point problem formulated in (4.14). We introduce finite dimensional subspaces V_h and M_h:

$$V_h \subset V, \quad M_h \subset M.$$

The Galerkin discretization of the problem (4.14) is as follows: find $(\boldsymbol{\phi}_h, \boldsymbol{\lambda}_h) \in V_h \times M_h$ such that

$$\hat{a}(\boldsymbol{\phi}_h, \boldsymbol{\psi}_h) + \hat{b}(\boldsymbol{\psi}_h, \boldsymbol{\lambda}_h) = f_1(\boldsymbol{\psi}_h) \quad \text{for all} \quad \boldsymbol{\psi}_h \in V_h \tag{4.20a}$$

$$\hat{b}(\boldsymbol{\phi}_h, \boldsymbol{\mu}_h) = f_2(\boldsymbol{\mu}_h) \quad \text{for all} \quad \boldsymbol{\mu}_h \in M_h. \tag{4.20b}$$

For the discretization error we have the following result, cf. [8, 20, 16].

Theorem 4.3.5 *Consider the variational problem* (4.14) *and its Galerkin discretization* (4.20)*, with continuous bilinear forms $\hat{a}(\cdot,\cdot)$ and $\hat{b}(\cdot,\cdot)$ that satisfy:*

$$\exists\ \beta > 0: \quad \sup_{\boldsymbol{\psi} \in V} \frac{\hat{b}(\boldsymbol{\psi}, \boldsymbol{\lambda})}{\|\boldsymbol{\psi}\|_V} \geq \beta \, \|\boldsymbol{\lambda}\|_M \qquad \forall\ \boldsymbol{\lambda} \in M, \tag{4.21a}$$

$$\exists\ \gamma > 0: \quad \hat{a}(\boldsymbol{\phi}, \boldsymbol{\phi}) \geq \gamma \, \|\boldsymbol{\phi}\|_V^2 \qquad \forall\ \boldsymbol{\phi} \in V, \tag{4.21b}$$

$$\exists\ \beta_h > 0: \quad \sup_{\boldsymbol{\psi}_h \in V_h} \frac{\hat{b}(\boldsymbol{\psi}_h, \boldsymbol{\lambda}_h)}{\|\boldsymbol{\psi}_h\|_V} \geq \beta_h \, \|\boldsymbol{\lambda}_h\|_M \quad \forall\ \boldsymbol{\lambda}_h \in M_h. \tag{4.21c}$$

Then the problem (4.14) *and its Galerkin discretization have unique solutions $(\boldsymbol{\phi}, \boldsymbol{\lambda})$ and $(\boldsymbol{\phi}_h, \boldsymbol{\lambda}_h)$, respectively. Furthermore the inequality*

$$\|\boldsymbol{\phi} - \boldsymbol{\phi}_h\|_V + \|\boldsymbol{\lambda} - \boldsymbol{\lambda}_h\|_M \leq C \Big(\inf_{\boldsymbol{\psi}_h \in V_h} \|\boldsymbol{\phi} - \boldsymbol{\psi}_h\|_V + \inf_{\boldsymbol{\mu}_h \in M_h} \|\boldsymbol{\lambda} - \boldsymbol{\mu}_h\|_M \Big)$$

holds, with $C = \sqrt{2}\big(1 + \gamma^{-1}\beta_h^{-2}(2\|\hat{a}\| + \|\hat{b}\|)^3\big)$.

Remark 4.3.6 The condition (4.21c) implies $\dim(V_h) \geq \dim(M_h)$. This can be shown by the following argument. Let $(\boldsymbol{\psi}_j)_{1\leq j\leq m}$ be a basis of V_h and $(\boldsymbol{\lambda}_i)_{1\leq i\leq k}$ a basis of M_h. Define the matrix $\mathbf{B} \in \mathbb{R}^{k\times m}$ by

$$\mathbf{B}_{ij} = \hat{b}(\boldsymbol{\psi}_j, \boldsymbol{\lambda}_i).$$

From (4.21c) it follows that for every $\boldsymbol{\lambda}_h \in M_h$, $\boldsymbol{\lambda}_h \neq 0$, there exists $\boldsymbol{\psi}_h \in V_h$ such that $\hat{b}(\boldsymbol{\psi}_h, \boldsymbol{\lambda}_h) \neq 0$. Thus for every $\mathbf{y} \in \mathbb{R}^k$, $\mathbf{y} \neq 0$, there exists $\mathbf{x} \in \mathbb{R}^m$ such that $\mathbf{y}^T\mathbf{B}\mathbf{x} \neq 0$, i.e., $\mathbf{x}^T\mathbf{B}^T\mathbf{y} \neq 0$. This implies that all columns of $\mathbf{B}^T$, and thus all rows of $\mathbf{B}$, are independent. A necessary condition for this is $k \leq m$. □

The first two conditions (4.21a) and (4.21b) are introduced in view of well-posedness of the given variational saddle point problem (4.14), cf. (4.17). The third condition (4.21c), which is called *discrete inf-sup condition*, is essential for the stability of the Galerkin discretization. Note that the constant C in the discretization error bound in Theorem 4.3.5 depends on β_h and that $C \to \infty$ if $\beta_h \downarrow 0$. The discrete inf-sup condition clearly depends on the specific pair of spaces (V_h, M_h) that is chosen.

References

1. R. A. Adams. *Sobolev Spaces*. Academic Press, 1975.
2. D. Arnold, F. Brezzi, B. Cockburn, and L. Marini. Unified analysis of discontinuous Galerkin methods for elliptic problems. *SIAM J. Numer. Anal.*, 39:1749–1779, 2002.
3. D. Boffi. Stability of higher-order triangular Hood-Taylor methods for the stationary Stokes equations. *Math. Models and Methods in Appl. Sci. (M3AS)*, 4:223–235, 1994.
4. D. Boffi. Three-dimensional finite element methods for the Stokes problem. *SIAM J. Numer. Anal.*, 34:664–670, 1997.
5. D. Braess. *Finite elements*. Cambridge University Press, Cambridge, second edition, 2001.
6. L. Brenner, S.and Scott. *The Mathematical Theory of Finite Element Methods*. Springer, New York, second edition, 2002.
7. F. Brezzi and R. Falk. Stability of higher-order Hood-Taylor methods. *SIAM J. Numer. Anal.*, 28:581–590, 1991.
8. F. Brezzi and M. Fortin. *Mixed and Hybrid Finite Element Methods*. Springer, Berlin, 1991.
9. M. Case, V. Ervin, A. Linke, and L. Rebholz. Improving mass conservation in FE approximations of the Navier-Stokes equations using continuous velocity fields: A connection between grad-div stabilization and Scott–Vogelius elements. Preprint 1510, Weierstrass Institute for Applied Analysis and Stochastics, 2010.
10. L. Cattabriga. Su un problema al contorno relativo al sistema di equazioni di Stokes. *Rend. Sem. Mat. Univ. Padov.*, 31:308–340, 1961.
11. P. Ciarlet. *The Finite Element Method for Elliptic Problems*. North-Holland, Amsterdam, 1978.
12. M. Crouzeix and P. Raviart. Conforming and non-conforming finite element methods for solving the stationary Stokes equations. *R.A.I.R.O. Anal. Numer.*, 7:33–76, 1973.
13. M. Dauge. Stationary Stokes and Navier-Stokes systems on two- or three-dimensional domains with corners. Part I: linearized equations. *SIAM J. Math. Anal.*, 20:74–97, 1989.
14. G. de Rham. *Variétés Differentiables*. Hermann, 1960.
15. G. Duvaut and J. Lions. *Les Inéquations en Mécanique et en Physique*. Dunod, Paris, 1972.
16. A. Ern and J.-L. Guermond. *Theory and Practice of Finite Elements*. Springer, New York, 2004.
17. M. Feistauer. *Mathematical Methods in Fluid Dynamics*. Longman Scientific & Technical, Harlow, 1993.
18. M. Feistauer, J. Felcman, and I. Straskraba. *Mathematical and Computational Methods for Compressible Flow*. Clarendon Press, Oxford, 2003.
19. L. Franca and S. Frey. Stabilized finite element methods: II. the incompressible Navier-Stokes equations. *Comp. Methods Appl. Mech. Engrg.*, 99:209–233, 1992.
20. V. Girault and P. Raviart. *Finite Element Methods for Navier-Stokes Equations*, volume 5 of *Springer Series in Computational Mathematics*. Springer, Berlin, 1986.
21. R. Glowinski. Numerical methods for fluids (part 3). In P. Ciarlet and J. Lions, editors, *Handbook of Numerical Analysis*, volume IX, pages 1–1083, Amsterdam, 2003. Elsevier.
22. M. Gurtin. *An Introduction to Continuum Mechanics*. Academic Press, New York, 1981.
23. E. Hairer and G. Wanner. *Solving Ordinary Differential Equations II: Stiff and Differential-algebraic Problems*. Springer, New York, 1991.
24. J. Hesthaven and T. Warburton. *Nodal Discontinuous Galerkin Methods: Algorithms, Analysis, and Applications*. Springer, New York, 2008.

25. 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.*, 23:750–777, 1986.
26. J. Heywood and R. Rannacher. Finite element approximation of the nonstationary Navier-Stokes problem, III. smoothing property and higher order estimates for spatial discretization. *SIAM J. Numer. Anal.*, 25:489–512, 1988.
27. 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.*, 27:353–384, 1990.
28. T. Hughes and L. Franca. A new finite element formulation for computational fluid dynamics: VII. the Stokes problem with various well-posed boundary conditions: symmetric formulations that converge for all velocity/pressure spaces. *Comp. Methods Appl. Mech. Engrg.*, 65:85–96, 1987.
29. D. Kay, P. Gresho, D. Griffiths, and D. Silvester. Adaptive time-stepping for incompressible flow; part II: Navier-Stokes equations. *SIAM J. Sci. Comput.*, 32:111–128, 2010.
30. O. Ladyzhenskaya. *Funktionalanalytische Untersuchungen der Navier-Stokesschen Gleichungen.* Akademie-Verlag, Berlin, 1965.
31. W. Layton. *Introduction to the Numerical Analysis of Incompressible Viscous Flows.* SIAM, Computational Science & Engineering. SIAM, Philadelphia, 2008.
32. M. Marion and R. Temam. Navier-Stokes equations: Theory and approximation. In P. Ciarlet and J. Lions, editors, *Handbook of Numerical Analysis*, volume VI, pages 503–689, Amsterdam, 1998. Elsevier.
33. J. Nečas. *Les Méthodes Directes en Théorie des Équations Elliptiques.* Masson, Paris, 1967.
34. A. Quarteroni and A. Valli. *Numerical Approximation of Partial Differential Equations*, volume 23 of *Springer Series in Computational Mathematics.* Springer-Verlag, Berlin, Heidelberg, 1994.
35. R. Rannacher. On the numerical solution of the incompressible Navier-Stokes equations. *Z. Angew. Math. Mech.*, 73(9):203–216, 1993.
36. C. Schwab. *p and hp−Finite Element Methods.* Clarendon Press, Oxford, 1998.
37. L. Scott and M. Vogelius. Norm estimates for a maximal right inverse of the divergence operator in spaces of piecewise polynomials. *RAIRO, Model. Math. Anal. Numer.*, 19:111–143, 1985.
38. G. Strang and G. Fix. *An Analysis of the Finite Element Method.* Prentice-Hall, Englewood Cliffs,NJ, 1973.
39. R. Temam. *Navier-Stokes Equations*, volume 2 of *Studies in Mathematics and its Applications.* North-Holland publishing company, Amsterdam, 1984.
40. R. Thatcher. Locally mass-conserving Taylor-Hood elements for two- and three-dimensional flow. *Int. J. Numer. Meth. Fluids*, 11(3):341–353, 1990.
41. L. Tobiska and R. Verfürth. Analysis of a streamline diffusion finite element method for the Stokes and Navier-Stokes equations. *SIAM J. Numer. Anal.*, 33:673–688, 1996.
42. S. Turek. *Efficient Solvers for Incompressible Flow Problems: An Algorithmic and Computational Approach*, volume 6 of *Lecture Notes in Computational Science and Engineering.* Springer, Berlin, Heidelberg, 1999.
43. H. Versteeg and W. Malalasekra. *An Introduction to Computational Fluid Dynamics: The Finite Volume Method, 2nd ed.* Prentice Hall, London, 2007.
44. P. Wesseling. *Principles of Computational Fluid Dynamics.* Springer, Berlin, 2000.
45. J. Wloka. *Partial Differential Equations.* Cambridge University Press, Cambridge, 1987.
46. E. Zeidler. *Nonlinear Functional Analysis and its Applications, II/A.* Springer, New York, 1990.
47. S. Zhang. A new family of stable mixed finite elements for the 3D Stokes equations. *Math. Comp.*, 74(250):543–554, 2005.
48. S. Zhang. On the $P1$ Powell-Sabin divergence-free finite element for the Stokes equations. *J. Comp. Math.*, 26:456–470, 2008.
49. S. Zhang. Quadratic divergence-free finite elements on Powell-Sabin tetrahedral grids. Preprint 10/2008,6/2009, Department of Mathematics, University of Delaware, 2009, www.math.udel.edu/ szhang/.