Computational Fluid Dynamics

Lecture Notes Summer Term 2018

R. Verfürth

Fakultät für Mathematik, Ruhr-Universität Bochum

Contents

CHAPTER I

Fundamentals

I.1. Modelization

I.1.1. Lagrangian and Eulerian representation. Consider a volume Ω occupied by a fluid which moves under the influence of interior and exterior forces. We denote by $\eta \in \Omega$ the position of an arbitrary particle at time $t = 0$ and by

$$x = \Phi(\eta, t)$$

its position at time $t > 0$. The basic assumptions are:

- $\Phi(\cdot, t) : \Omega \to \Omega$ is for all times $t \geq 0$ a bijective differentiable mapping,
- its inverse mapping is differentiable,
- the transformation $\Phi(\cdot, t)$ is orientation-preserving,
- $\Phi(\cdot, 0)$ is the identity.

There are two ways to look at the fluid flow:

(1) We fix η and look at the trajectory $t \mapsto \Phi(\eta, t)$. This is the *Lagrangian representation*. Correspondingly η is called *Lagrangian coordinate*. The Langrangian coordinate system moves with the fluid.

(2) We fix the point x and look at the trajectory $t \mapsto \Phi(\cdot, t)^{-1}(x)$ which passes through x. This is the *Eulerian representation*. Correspondingly x is called *Eulerian coordinate*. The Eulerian coordinate system is fixed.

I.1.2. Velocity. We denote by

$$D\Phi = \Big(\frac{\partial \Phi_i}{\partial \eta_j}\Big)_{1 \leq i,j \leq 3}$$

the Jacobian matrix of $\Phi(\cdot, t)$ and by

$$J = \det D\Phi$$

its Jacobian determinant. The preservation of orientation is equivalent to

$$J(\eta, t) > 0 \quad \text{for all } t > 0 \text{ and all } \eta \in \Omega.$$

The *velocity* of the flow at point $x = \Phi(\eta, t)$ is defined by

$$\boxed{\mathbf{v}(x,t) = \frac{\partial}{\partial t}\Phi(\eta,t), \quad x = \Phi(\eta,t).}$$

I.1.3. Transport theorem. Consider an arbitrary volume $V \subset \Omega$ and denote by

$$V(t) = \Phi(\cdot,t)(V) = \{\Phi(\eta,t) : \eta \in V\}$$

its shape at time $t > 0$. Then the following *transport theorem* holds for all differentiable mappings $f : \Omega \times (0,\infty) \to \mathbb{R}$

$$\boxed{\frac{d}{dt}\int_{V(t)} f(x,t)\mathrm{d}x = \int_{V(t)} \Big\{\frac{\partial f}{\partial t}(x,t) + \operatorname{div}(f\mathbf{v})(x,t)\Big\}\mathrm{d}x.}$$

Example I.1.1. We consider a one-dimensional model situation, i.e. $\Omega = (a,b)$ is an interval and x and η are real numbers. Then $V = (\alpha,\beta)$ and $V(t) = (\alpha(t),\beta(t))$ are intervals, too. The transformation rule for integrals gives

$$\begin{aligned}\int_{V(t)} f(x,t)\mathrm{d}x &= \int_{\alpha(t)}^{\beta(t)} f(x,t)\mathrm{d}x \\ &= \int_\alpha^\beta f(\Phi(\eta,t),t)\frac{\partial\Phi}{\partial\eta}(\eta,t)\mathrm{d}\eta.\end{aligned}$$

Differentiating this equality and using the transformation rule we get

$$\begin{aligned}\frac{d}{dt}\int_{V(t)} f(x,t)\mathrm{d}x &= \int_\alpha^\beta \frac{d}{dt}\left\{f(\Phi(\eta,t),t)\frac{\partial\Phi}{\partial\eta}(\eta,t)\right\}\mathrm{d}\eta \\ &= \underbrace{\int_\alpha^\beta \frac{\partial}{\partial t}f(\Phi(\eta,t),t)\frac{\partial\Phi}{\partial\eta}(\eta,t)\mathrm{d}\eta}_{=\int_{V(t)}\frac{\partial}{\partial t}f(x,t)\ \mathrm{d}x} \\ &\quad + \underbrace{\int_\alpha^\beta \frac{\partial}{\partial x}f(\Phi(\eta,t),t)\ \underbrace{\frac{\partial\Phi}{\partial t}(\eta,t)}_{=\mathbf{v}(\Phi(\eta,t),t)}\ \frac{\partial\Phi}{\partial\eta}(\eta,t)\mathrm{d}\eta}_{=\int_{V(t)}\frac{\partial}{\partial x}f(x,t)\mathbf{v}(x,t)\mathrm{d}x} \\ &\quad + \underbrace{\int_\alpha^\beta f(\Phi(\eta,t),t)\ \underbrace{\frac{\partial}{\partial t}\frac{\partial}{\partial\eta}\Phi(\eta,t)}_{=\frac{\partial}{\partial\eta}\mathbf{v}(\Phi(\eta,t),t)=\frac{\partial}{\partial x}\mathbf{v}(\Phi(\eta,t),t)\frac{\partial\Phi}{\partial\eta}(\eta,t)}\ \mathrm{d}\eta}_{=\int_{V(t)} f(x,t)\frac{\partial}{\partial x}\mathbf{v}(x,t)\mathrm{d}x}\end{aligned}$$

$$
\begin{aligned}
&= \int_{V(t)} \frac{\partial}{\partial t} f(x,t) \mathrm{d}x \\
&\quad + \int_{V(t)} \frac{\partial}{\partial x} f(x,t) \mathbf{v}(x,t) \mathrm{d}x \\
&\quad + \int_{V(t)} f(x,t) \frac{\partial}{\partial x} \mathbf{v}(x,t) \mathrm{d}x \\
&= \int_{V(t)} \frac{\partial}{\partial t} f(x,t) + \frac{\partial}{\partial x} \left\{ f(x,t) \mathbf{v}(x,t) \right\} \mathrm{d}x.
\end{aligned}
$$

This proves the transport theorem in one dimension.

I.1.4. Conservation of mass. We denote by $\rho(x,t)$ the density of the fluid. Then

$$\int_{V(t)} \rho(x,t) \mathrm{d}x$$

is the total mass of the volume $V(t)$. The conservation of mass and the transport theorem therefore imply that

$$
\begin{aligned}
0 &= \frac{d}{dt} \int_{V(t)} \rho(x,t) \mathrm{d}x \\
&= \int_{V(t)} \Big\{ \frac{\partial \rho}{\partial t}(x,t) + \operatorname{div}(\rho \mathbf{v})(x,t) \Big\} \mathrm{d}x.
\end{aligned}
$$

This gives the equation for the *conservation of mass*:

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

I.1.5. Cauchy theorem. Fluid and continuum mechanics are based on three fundamental assumptions concerning the interior forces:

- interior forces act via the surface of a volume $V(t)$,
- interior forces only depend on the normal direction of the surface of the volume,
- interior forces are additive and continuous.

Due to the *Cauchy theorem* these assumptions imply that the interior forces acting on a volume $V(t)$ must be of the form

$$\int_{\partial V(t)} \underline{\mathbf{T}} \cdot \mathbf{n} \mathrm{d}S$$

with a tensor $\underline{\mathbf{T}} : \Omega \to \mathbb{R}^{3\times 3}$. Here, as usual, $\partial V(t)$ denotes the boundary of the volume $V(t)$, $\mathbf{n}$ is the unit outward normal, and dS denotes

the surface element.
The integral theorem of Gauß then yields

$$\int_{\partial V(t)} \underline{\mathbf{T}} \cdot \mathbf{n} \mathrm{d}S = \int_{V(t)} \operatorname{div} \underline{\mathbf{T}} \mathrm{d}x.$$

I.1.6. Conservation of momentum. The total momentum of a volume $V(t)$ in the fluid is given by

$$\int_{V(t)} \rho(x,t)\mathbf{v}(x,t)\mathrm{d}x.$$

Due to the transport theorem its temporal variation equals

$$\begin{aligned}
&\frac{d}{dt}\int_{V(t)} \rho(x,t)\mathbf{v}(x,t)\mathrm{d}x \\
&\quad = \Big(\frac{d}{dt}\int_{V(t)} \rho(x,t)v_i(x,t)\mathrm{d}x\Big)_{1\le i\le 3} \\
&\quad = \Big(\int_{V(t)} \Big\{\frac{\partial}{\partial t}(\rho v_i)(x,t) + \operatorname{div}(\rho v_i \mathbf{v})(x,t)\Big\}\mathrm{d}x\Big)_{1\le i\le 3} \\
&\quad = \int_{V(t)} \Big\{\frac{\partial}{\partial t}(\rho \mathbf{v})(x,t) + \operatorname{div}(\rho \mathbf{v}\otimes\mathbf{v})(x,t)\Big\}\mathrm{d}x
\end{aligned}$$

where

$$\mathbf{v}\otimes\mathbf{u} = (v_i u_j)_{1\le i,j\le 3}.$$

Due to the conservation of momentum this quantity must be balanced by exterior and interior forces. Exterior forces must be of the form

$$\int_{V(t)} \rho\mathbf{f}\mathrm{d}x.$$

The Cauchy theorem tells that the interior forces are of the form

$$\int_{V(t)} \operatorname{div} \underline{\mathbf{T}} \mathrm{d}x.$$

Hence the conservation of momentum takes the integral form

$$\int_{V(t)} \Big\{\frac{\partial}{\partial t}(\rho\mathbf{v}) + \operatorname{div}(\rho\mathbf{v}\otimes\mathbf{v})\Big\}\mathrm{d}x = \int_{V(t)} \big\{\rho\mathbf{f} + \operatorname{div}\underline{\mathbf{T}}\big\}\mathrm{d}x.$$

This gives the equation for the *conservation of momentum*:

$$\boxed{\frac{\partial}{\partial t}(\rho\mathbf{v}) + \operatorname{div}(\rho\mathbf{v}\otimes\mathbf{v}) = \rho\mathbf{f} + \operatorname{div}\underline{\mathbf{T}} \quad \text{in } \Omega\times(0,\infty).}$$

I.1.7. Conservation of energy. We denote by e the *total energy*. The transport theorem implies that its temporal variation is given by

$$\frac{d}{dt}\int_{V(t)} e\mathrm{d}x = \int_{V(t)} \Big\{\frac{\partial e}{\partial t}(x,t) + \operatorname{div}(e\mathbf{v})(x,t)\Big\}\mathrm{d}x.$$

Due to the conservation of energy this quantity must be balanced by the energies due to the exterior and interior forces and by the change of the internal energy.
The contributions of the exterior and interior forces are given by

$$\int_{V(t)} \rho\mathbf{f}\cdot\mathbf{v}\mathrm{d}x$$

and

$$\int_{\partial V(t)} \mathbf{n}\cdot\underline{\mathbf{T}}\cdot\mathbf{v}\mathrm{d}S = \int_{V(t)} \operatorname{div}(\underline{\mathbf{T}}\cdot\mathbf{v})\mathrm{d}x.$$

Due to the Cauchy theorem and the integral theorem of Gauß, the change of internal energy must be of the form

$$\int_{\partial V(t)} \boldsymbol{\sigma}\cdot\mathbf{n}\mathrm{d}S = \int_{V(t)} \operatorname{div}\boldsymbol{\sigma}\mathrm{d}x$$

with a vector-field $\boldsymbol{\sigma}:\Omega\to\mathbb{R}^3$.
Hence the conservation of energy takes the integral form

$$\int_{V(t)} \Big\{\frac{\partial}{\partial t}e + \operatorname{div}(e\mathbf{v})\Big\}\mathrm{d}x = \int_{V(t)} \big\{\rho\mathbf{f}\cdot\mathbf{v} + \operatorname{div}(\underline{\mathbf{T}}\cdot\mathbf{v}) + \operatorname{div}\boldsymbol{\sigma}\big\}\mathrm{d}x.$$

This gives the equation for the *conservation of energy*:

$$\frac{\partial}{\partial t}e + \operatorname{div}(e\mathbf{v}) = \rho\mathbf{f}\cdot\mathbf{v} + \operatorname{div}(\underline{\mathbf{T}}\cdot\mathbf{v}) + \operatorname{div}\boldsymbol{\sigma} \quad \text{in } \Omega\times(0,\infty).$$

I.1.8. Constitutive laws. The equations for the conservation of mass, momentum and energy must be complemented by constitutive laws. These are based on the following fundamental assumptions:

- $\underline{\mathbf{T}}$ only depends on the gradient of the velocity.
- The dependence on the velocity gradient is linear.
- $\underline{\mathbf{T}}$ is symmetric. (Due to the Cauchy theorem this is a consequence of the conservation of angular momentum.)
- In the absence of internal friction, $\underline{\mathbf{T}}$ is diagonal and proportional to the pressure, i.e. all interior forces act in normal direction.
- $e = \rho\varepsilon + \frac{1}{2}\rho|\mathbf{v}|^2$. ($\varepsilon$ is called *internal energy* and is often identified with the temperature.)

- $\boldsymbol{\sigma}$ is proportional to the variation of the internal energy, i.e. $\boldsymbol{\sigma} = \alpha \nabla \varepsilon$.

The conditions on $\underline{\mathbf{T}}$ imply that it must be of the form

$$\underline{\mathbf{T}} = 2\lambda \underline{\mathbf{D}}(\mathbf{v}) + \mu(\operatorname{div} \mathbf{v})\, \underline{\mathbf{I}} - p\underline{\mathbf{I}}.$$

Here,

$$\underline{\mathbf{I}} = (\delta_{ij})_{1 \le i,j \le 3}$$

denotes the *unit tensor*,

$$\underline{\mathbf{D}}(\mathbf{v}) = \frac{1}{2}(\nabla \mathbf{v} + \nabla \mathbf{v}^t) = \frac{1}{2}\Big(\frac{\partial v_i}{\partial x_j} + \frac{\partial v_j}{\partial x_i}\Big)_{1 \le i,j \le 3}$$

is the *deformation tensor*,

$$p = p(\rho, \varepsilon)$$

denotes the *pressure* and $\lambda, \mu \in \mathbb{R}$ are the *dynamic viscosities* of the fluid.
The equation for the pressure is also called *equation of state*. For an ideal gas, e.g., it takes the form $p(\rho, \varepsilon) = (\gamma - 1)\rho\varepsilon$ with $\gamma > 1$.
Note that $\operatorname{div} \mathbf{v}$ is the trace of the deformation tensor $\underline{\mathbf{D}}(\mathbf{v})$.

EXAMPLE I.1.2. The physical dimension of the dynamic viscosities λ and μ is $\mathtt{kg\ m^{-1}s^{-1}}$ and is called *Poise*; the one of the density ρ is $\mathtt{kg\ m^{-3}}$. In §I.1.12 (p. 15) we will introduce the kinematic viscosity $\nu = \frac{\lambda}{\rho}$. Its physical dimension is $\mathtt{m^2\ s^{-1}}$ and is called *Stokes*. Table I.1.1 gives some relevant values of these quantities.

TABLE I.1.1. Viscosities of some fluids and gazes

	λ	ρ	ν
Water 20° C	$1.005\ 10^{-3}$	1000	$1.005\ 10^{-6}$
Alcohol 20° C	$1.19\ 10^{-3}$	790	$1.506\ 10^{-6}$
Ether 20° C	$2.43\ 10^{-5}$	716	$3.394\ 10^{-8}$
Glycerine 20° C	1.499	1260	$1.190\ 10^{-3}$
Air 0° C 1 atm	$1.71\ 10^{-5}$	1.293	$1.322\ 10^{-5}$
Hydrogen 0° C	$8.4\ 10^{-6}$	$8.99\ 10^{-2}$	$9.344\ 10^{-5}$

I.1.9. Compressible Navier-Stokes equations in conservative form. Collecting all results we obtain the *compressible Navier-Stokes equations in conservative form*:

$$
\begin{aligned}
\frac{\partial \rho}{\partial t} + \operatorname{div}(\rho \mathbf{v}) &= 0 \\
\frac{\partial (\rho \mathbf{v})}{\partial t} + \operatorname{div}(\rho \mathbf{v} \otimes \mathbf{v}) &= \rho \mathbf{f} + 2\lambda \operatorname{div} \underline{\mathbf{D}}(\mathbf{v}) \\
&\quad + \mu \operatorname{grad} \operatorname{div} \mathbf{v} - \operatorname{grad} p \\
&= \rho \mathbf{f} + \lambda \Delta \mathbf{v} \\
&\quad + (\lambda + \mu) \operatorname{grad} \operatorname{div} \mathbf{v} - \operatorname{grad} p \\
\frac{\partial e}{\partial t} + \operatorname{div}(e \mathbf{v}) &= \rho \mathbf{f} \cdot \mathbf{v} + 2\lambda \operatorname{div}[\underline{\mathbf{D}}(\mathbf{v}) \cdot \mathbf{v}] \\
&\quad + \mu \operatorname{div}[\operatorname{div} \mathbf{v} \cdot \mathbf{v}] \\
&\quad - \operatorname{div}(p \mathbf{v}) + \alpha \Delta \varepsilon \\
&= \{\rho \mathbf{f} + 2\lambda \operatorname{div} \underline{\mathbf{D}}(\mathbf{v}) + \mu \operatorname{grad} \operatorname{div} \mathbf{v} \\
&\quad - \operatorname{grad} p\} \cdot \mathbf{v} \\
&\quad + \lambda \underline{\mathbf{D}}(\mathbf{v}) : \underline{\mathbf{D}}(\mathbf{v}) + \mu (\operatorname{div} \mathbf{v})^2 \\
&\quad - p \operatorname{div} \mathbf{v} + \alpha \Delta \varepsilon \\
p &= p(\rho, \varepsilon) \\
e &= \rho \varepsilon + \frac{1}{2} \rho |\mathbf{v}|^2 .
\end{aligned}
$$

I.1.10. Euler equations. In the inviscid case, i.e. $\lambda = \mu = 0$, the compressible Navier-Stokes equations in conservative form reduce to the so-called *Euler equations* for an ideal gas:

$$
\begin{aligned}
\frac{\partial \rho}{\partial t} + \operatorname{div}(\rho \mathbf{v}) &= 0 \\
\frac{\partial (\rho \mathbf{v})}{\partial t} + \operatorname{div}(\rho \mathbf{v} \otimes \mathbf{v} + p \underline{\mathbf{I}}) &= \rho \mathbf{f} \\
\frac{\partial e}{\partial t} + \operatorname{div}(e \mathbf{v} + p \mathbf{v}) &= \rho \mathbf{f} \cdot \mathbf{v} + \alpha \Delta \varepsilon \\
p &= p(\rho, \varepsilon) \\
e &= \rho \varepsilon + \frac{1}{2} \rho |\mathbf{v}|^2 .
\end{aligned}
$$

I.1.11. Compressible Navier-Stokes equations in non-conservative form. Inserting the first equation of §I.1.9 in the left-hand

side of the second equation yields

$$\begin{aligned}
&\frac{\partial(\rho\mathbf{v})}{\partial t} + \operatorname{div}(\rho\mathbf{v}\otimes\mathbf{v}) \\
&\quad = \left(\frac{\partial\rho}{\partial t}\right)\mathbf{v} + \rho\frac{\partial\mathbf{v}}{\partial t} + (\operatorname{div}(\rho\mathbf{v}))\mathbf{v} + \rho(\mathbf{v}\cdot\nabla)\mathbf{v} \\
&\quad = \rho\left[\frac{\partial\mathbf{v}}{\partial t} + (\mathbf{v}\cdot\nabla)\mathbf{v}\right].
\end{aligned}$$

Inserting the first two equations of §I.1.9 in the left-hand side of the third equation implies

$$\begin{aligned}
&\lambda\underline{\mathbf{D}}(\mathbf{v}) : \underline{\mathbf{D}}(\mathbf{v}) + \mu(\operatorname{div}\mathbf{v})^2 - p\operatorname{div}\mathbf{v} + \alpha\Delta\varepsilon \\
&\quad = -\{\rho\mathbf{f} + 2\lambda\operatorname{div}\underline{\mathbf{D}}(\mathbf{v}) + \mu\operatorname{grad}\operatorname{div}\mathbf{v} - \operatorname{grad}p\}\cdot\mathbf{v} \\
&\qquad + \frac{\partial(\rho\varepsilon + \frac{1}{2}\rho|\mathbf{v}|^2)}{\partial t} + \operatorname{div}\left(\rho\varepsilon\mathbf{v} + \frac{1}{2}\rho|\mathbf{v}|^2\mathbf{v}\right) \\
&\quad = -\left\{\frac{\partial(\rho\mathbf{v})}{\partial t} + \operatorname{div}(\rho\mathbf{v}\otimes\mathbf{v})\right\}\cdot\mathbf{v} \\
&\qquad + \varepsilon\frac{\partial\rho}{\partial t} + \varepsilon\operatorname{div}(\rho\mathbf{v}) + \rho\frac{\partial\varepsilon}{\partial t} + \rho\mathbf{v}\cdot\operatorname{grad}\varepsilon \\
&\qquad + \frac{1}{2}|\mathbf{v}|^2\frac{\partial\rho}{\partial t} + \frac{1}{2}|\mathbf{v}|^2\operatorname{div}(\rho\mathbf{v}) + \rho\frac{1}{2}\frac{\partial|\mathbf{v}|^2}{\partial t} + \rho\mathbf{v}\cdot\operatorname{grad}\left(\frac{1}{2}|\mathbf{v}|^2\right) \\
&\quad = -\rho\mathbf{v}\cdot\left[\frac{\partial\mathbf{v}}{\partial t} + (\mathbf{v}\cdot\nabla)\mathbf{v}\right] \\
&\qquad + \rho\frac{\partial\varepsilon}{\partial t} + \rho\mathbf{v}\cdot\operatorname{grad}\varepsilon \\
&\qquad + \rho\mathbf{v}\cdot\frac{\partial\mathbf{v}}{\partial t} + \rho\mathbf{v}\cdot[(\mathbf{v}\cdot\nabla)\mathbf{v}] \\
&\quad = \rho\frac{\partial\varepsilon}{\partial t} + \rho\mathbf{v}\cdot\operatorname{grad}\varepsilon.
\end{aligned}$$

With these simplifications we obtain the *compressible Navier-Stokes equations in non-conservative form*

$$\begin{aligned}
\frac{\partial\rho}{\partial t} + \operatorname{div}(\rho\mathbf{v}) &= 0 \\
\rho\left[\frac{\partial\mathbf{v}}{\partial t} + (\mathbf{v}\cdot\nabla)\mathbf{v}\right] &= \rho\mathbf{f} + \lambda\Delta\mathbf{v} + (\lambda+\mu)\operatorname{grad}\operatorname{div}\mathbf{v} \\
&\quad - \operatorname{grad}p \\
\rho\left[\frac{\partial\varepsilon}{\partial t} + \rho\mathbf{v}\cdot\operatorname{grad}\varepsilon\right] &= \lambda\underline{\mathbf{D}}(\mathbf{v}) : \underline{\mathbf{D}}(\mathbf{v}) + \mu(\operatorname{div}\mathbf{v})^2 - p\operatorname{div}\mathbf{v} \\
&\quad + \alpha\Delta\varepsilon \\
p &= p(\rho,\varepsilon).
\end{aligned}$$

I.1.12. Instationary incompressible Navier-Stokes equations. Next we assume that the density ρ is constant, replace p by $\frac{p}{\rho}$, denote by

$$\boxed{\nu = \frac{\lambda}{\rho}}$$

the *kinematic viscosity*, and suppress the third equation in §I.1.11 (conservation of energy). This yields the *instationary incompressible Navier-Stokes equations*:

$$\boxed{\begin{aligned} \operatorname{div} \mathbf{v} &= 0 \\ \frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v} \cdot \nabla)\mathbf{v} &= \mathbf{f} + \nu \Delta \mathbf{v} - \operatorname{grad} p. \end{aligned}}$$

REMARK I.1.3. We say that a fluid is *incompressible* if the volume of any sub-domain $V \subset \Omega$ remains constant for all times $t > 0$, i.e.

$$\int_{V(t)} \mathrm{d}x = \int_V \mathrm{d}x \quad \text{for all } t > 0.$$

The transport theorem then implies that

$$0 = \frac{d}{dt} \int_{V(t)} \mathrm{d}x = \int_{V(t)} \operatorname{div} \mathbf{v} \mathrm{d}x.$$

Hence, incompressibility is equivalent to the equation

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

REMARK I.1.4. The incompressible Navier-Stokes equations are often re-scaled to a non-dimensional form. To this end we introduce a reference length L, a reference time T, a reference velocity U, a reference pressure P, and a reference force F and introduce new variables and quantities by $x = Ly$, $t = T\tau$, $\mathbf{v} = U\mathbf{u}$, $p = Pq$, $\mathbf{f} = F\mathbf{g}$. Physical reasons suggest to fix $T = \frac{L}{U}$. When re-writing the momentum equation in the new quantities we obtain

$$\begin{aligned} F\mathbf{g} = \mathbf{f} &= \frac{U}{T}\frac{\partial \mathbf{u}}{\partial \tau} + \frac{U^2}{L}(\mathbf{u} \cdot \nabla_y)\mathbf{u} - \frac{\nu U}{L^2}\Delta_y \mathbf{u} + \frac{P}{L}\nabla_y q \\ &= \frac{\nu U}{L^2}\left\{ \frac{L^2}{\nu T}\frac{\partial \mathbf{u}}{\partial \tau} + \frac{LU}{\nu}(\mathbf{u} \cdot \nabla_y)\mathbf{u} - \Delta_y \mathbf{u} + \frac{PL}{\nu U}\nabla_y q \right\}. \end{aligned}$$

The physical dimension of F and $\frac{\nu U}{L^2}$ is $\mathtt{m}\ \mathtt{s}^{-2}$. The quantities $\frac{L^2}{\nu T}$, $\frac{PL}{\nu U}$, and $\frac{LU}{\nu}$ are dimensionless. The quantities L, U, and ν are determined by the physical problem. The quantities F and P are fixed by the conditions $F = \frac{\nu U}{L^2}$ and $\frac{PL}{\nu U} = 1$. Thus there remains only the dimensionless parameter

$$\boxed{Re = \frac{LU}{\nu}.}$$

It is called *Reynolds' number* and is a measure for the complexity of the flow. It was introduced in 1883 by *Osborne Reynolds* (1842 – 1912).

EXAMPLE I.1.5. Consider an airplane at cruising speed and an oil-vessel. The relevant quantities for the airplane are

$$\begin{aligned} U &\approx 900\mathtt{km\ h}^{-1} \approx 250\mathtt{m\ s}^{-1} \\ L &\approx 50\mathtt{m} \\ \nu &\approx 1.322\ 10^{-5}\mathtt{m}^2\ \mathtt{s}^{-1} \\ Re &\approx 9.455\ 10^8. \end{aligned}$$

The corresponding quantities for the oil-vessel are

$$\begin{aligned} U &\approx 20\mathtt{knots} \approx 36\mathtt{km\ h}^{-1} \approx 10\mathtt{m\ s}^{-1} \\ L &\approx 300\mathtt{m} \\ \nu &\approx 1.005\ 10^{-6}\mathtt{m}^2\ \mathtt{s}^{-1} \\ Re &\approx 2.985\ 10^9. \end{aligned}$$

I.1.13. Stationary incompressible Navier-Stokes equations. As a next step of simplification, we assume that we are in a stationary regime, i.e. all temporal derivatives vanish. This yields the *stationary incompressible Navier-Stokes equations*:

$$\boxed{\begin{aligned} \operatorname{div}\mathbf{v} &= 0 \\ -\nu\Delta\mathbf{v} + (\mathbf{v}\cdot\nabla)\mathbf{v} + \operatorname{grad} p &= \mathbf{f}. \end{aligned}}$$

I.1.14. Stokes equations. In a last step we linearize the Navier-Stokes equations at velocity $\mathbf{v} = 0$ and re-scale the viscosity ν to 1. This gives the *Stokes equations*:

$$\boxed{\begin{aligned} \operatorname{div}\mathbf{v} &= 0 \\ -\Delta\mathbf{v} + \operatorname{grad} p &= \mathbf{f}. \end{aligned}}$$

REMARK I.1.6. All CFD algorithms solve complicated flow problems via a nested sequence of simplification steps similar to the described procedure. Therefore it is mandatory to dispose of efficient discretization schemes and solvers for simplified problems as, e.g., the Stokes equations, which are at the heart in the inmost loop.

I.1.15. Initial and boundary conditions. The equations of §§ I.1.9 – I.1.12 are time-dependent. They must be complemented by initial conditions for the velocity $\mathbf{v}$, the internal energy ε and – for the problems of §§I.1.9 – I.1.11 – the density ρ.

The Euler equations of §I.1.10 (p. 13) are hyperbolic and require boundary conditions on the inflow-boundary.

The problems of §§I.1.9 and I.1.11 – I.1.14 are of second order in space and require boundary conditions everywhere on the boundary $\Gamma = \partial\Omega$. For the energy-equations in §§I.1.9 (p. 12) and I.1.11 (p. 13) one may choose either Dirichlet (prescribed energy) or Neumann (prescribed energy flux) boundary conditions. A mixture of both conditions is also possible.

The situation is less evident for the mass and momentum equations in §§I.1.9 and I.1.11 – I.1.14. Around 1845, Sir *George Gabriel Stokes* (1819 – 1903) suggested that – due to the friction – the fluid will adhere at the boundary. This leads to the Dirichlet boundary condition

$$\mathbf{v} = 0 \quad \text{on } \Gamma.$$

More generally one can impose the condition $\mathbf{v} = \mathbf{v}_\Gamma$ with a given boundary velocity $\mathbf{v}_\Gamma$.

Around 1827, *Pierre Louis Marie Henri Navier* (1785 – 1836) had suggested the more general boundary condition

$$\begin{aligned} \lambda_n \mathbf{v}\cdot\mathbf{n} + (1-\lambda_n)\mathbf{n}\cdot\underline{\mathbf{T}}\cdot\mathbf{n} &= 0 \\ \lambda_t[\mathbf{v} - (\mathbf{v}\cdot\mathbf{n})\mathbf{n}] + (1-\lambda_t)[\underline{\mathbf{T}}\cdot\mathbf{n} - (\mathbf{n}\cdot\underline{\mathbf{T}}\cdot\mathbf{n})\mathbf{n}] &= 0 \end{aligned}$$

on Γ with parameters $\lambda_n, \lambda_t \in [0,1]$ that depend on the actual flow-problem. More generally one may replace the homogeneous right-hand sides by given known functions.

The first equation of Navier refers to the normal components of $\mathbf{v}$ and $\mathbf{n}\cdot\underline{\mathbf{T}}$, the second equation refers to the tangential components of $\mathbf{v}$ and $\mathbf{n}\cdot\underline{\mathbf{T}}$. The special case $\lambda_n = \lambda_t = 1$ obviously yields the *no-slip condition* of Stokes. The case $\lambda_n = 1, \lambda_t = 0$ on the other hand gives the *slip condition*

$$\begin{aligned} \mathbf{v}\cdot\mathbf{n} &= 0 \\ \underline{\mathbf{T}}\cdot\mathbf{n} - (\mathbf{n}\cdot\underline{\mathbf{T}}\cdot\mathbf{n})\mathbf{n} &= 0. \end{aligned}$$

The question, which boundary condition is a better description of the reality, was decided in the 19th century by experiments with pendulums submerged into a viscous fluid. They were in favour of the no-slip condition of Stokes. The viscosities of the fluids, however, were similar to that of honey and the velocities were only a few meters per second. When dealing with much higher velocities or much smaller viscosities, the slip condition of Navier is more appropriate. A particular example for this situation is the re-entry of a space vehicle. Other examples are

coating problems where the location of the boundary is an unknown and is determined by the interaction of capillary and viscous forces.

I.2. Notations and basic results

I.2.1. Domains and functions. The following notations concerning domains and functions will frequently be used:

Ω open, bounded, connected set in $\mathbb{R}^n$, $n \in \{2, 3\}$;
Γ boundary of Ω supposed to be Lipschitz-continuous;
$\mathbf{n}$ exterior unit normal to Ω;
$p, q, r, \ldots$ scalar functions with values in $\mathbb{R}$;
$\mathbf{u}, \mathbf{v}, \mathbf{w}, \ldots$ vector-fields with values in $\mathbb{R}^n$;
$\underline{\mathbf{S}}, \underline{\mathbf{T}}, \ldots$ tensor-fields with values in $\mathbb{R}^{n\times n}$;
$\underline{\mathbf{I}}$ unit tensor;
∇ gradient;
div divergence;

$$\operatorname{div}\mathbf{u} = \sum_{i=1}^{n} \frac{\partial u_i}{\partial x_i};$$

$$\operatorname{div}\underline{\mathbf{T}} = \left(\sum_{i=1}^{n} \frac{\partial T_{ij}}{\partial x_i} \right)_{1\le j\le n};$$

$\Delta = \operatorname{div}\nabla$ Laplace operator;

$$\underline{\mathbf{D}}(\mathbf{u}) = \frac{1}{2}\left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} \right)_{1\le i,j\le n}$$ deformation tensor;

$\mathbf{u}\cdot\mathbf{v}$ inner product;
$\underline{\mathbf{S}} : \underline{\mathbf{T}}$ dyadic product (inner product of tensors);
$\mathbf{u}\otimes\mathbf{v} = (\mathbf{u}_i\mathbf{v}_j)_{1\le i,j\le n}$ tensorial product.

I.2.2. Differentiation of products. The product formula for differentiation yields the following formulae for the differentiation of products of scalar functions, vector-fields and tensor-fields:

$$\begin{aligned} \operatorname{div}(p\mathbf{u}) &= \nabla p\cdot\mathbf{u} + p\operatorname{div}\mathbf{u}, \\ \operatorname{div}(\underline{\mathbf{T}}\cdot\mathbf{u}) &= (\operatorname{div}\underline{\mathbf{T}})\cdot\mathbf{u} + \underline{\mathbf{T}} : \underline{\mathbf{D}}(\mathbf{u}), \\ \operatorname{div}(\mathbf{u}\otimes\mathbf{v}) &= (\operatorname{div}\mathbf{u})\mathbf{v} + (\mathbf{u}\cdot\nabla)\mathbf{v}. \end{aligned}$$

I.2.3. Integration by parts formulae. The above product formulae and the Gauß theorem for integrals give rise to the following integration by parts formulae:

$$\int_\Gamma p\mathbf{u}\cdot\mathbf{n}\mathrm{d}S = \int_\Omega \nabla p\cdot\mathbf{u}\mathrm{d}x + \int_\Omega p\,\mathrm{div}\,\mathbf{u}\mathrm{d}x,$$
$$\int_\Gamma \mathbf{n}\cdot\underline{\mathbf{T}}\cdot\mathbf{u}\mathrm{d}S = \int_\Omega (\mathrm{div}\,\underline{\mathbf{T}})\cdot\mathbf{u}\mathrm{d}x + \int_\Omega \underline{\mathbf{T}}:\underline{\mathbf{D}}(\mathbf{u})\mathrm{d}x,$$
$$\int_\Gamma \mathbf{n}\cdot(\mathbf{u}\otimes\mathbf{v})\mathrm{d}S = \int_\Omega (\mathrm{div}\,\mathbf{u})\mathbf{v}\mathrm{d}x + \int_\Omega (\mathbf{u}\cdot\nabla)\mathbf{v}\mathrm{d}x.$$

I.2.4. Weak derivatives. Recall that $\overline{A}$ denotes the closure of a set $A\subset\mathbb{R}^n$.

Example I.2.1. For the sets

$A=\{x\in\mathbb{R}^3 : x_1^2+x_2^2+x_3^2<1\}$	open unit ball
$B=\{x\in\mathbb{R}^3 : 0<x_1^2+x_2^2+x_3^2<1\}$	punctuated open unit ball
$C=\{x\in\mathbb{R}^3 : 1<x_1^2+x_2^2+x_3^2<2\}$	open annulus

we have

$\overline{A}=\{x\in\mathbb{R}^3 : x_1^2+x_2^2+x_3^2\le 1\}$	closed unit ball
$\overline{B}=\{x\in\mathbb{R}^3 : x_1^2+x_2^2+x_3^2\le 1\}$	closed unit ball
$\overline{C}=\{x\in\mathbb{R}^3 : 1\le x_1^2+x_2^2+x_3^2\le 2\}$	closed annulus.

Given a continuous function $\varphi:\mathbb{R}^n\to\mathbb{R}$, we denote its *support* by

$$\mathrm{supp}\,\varphi = \overline{\{x\in\mathbb{R}^n : \varphi(x)\neq 0\}}.$$

The set of all functions that are infinitely differentiable and have their support contained in Ω is denoted by $C_0^\infty(\Omega)$:

$$C_0^\infty(\Omega) = \{\varphi\in C^\infty(\Omega) : \mathrm{supp}\,\varphi\subset\Omega\}.$$

Remark I.2.2. The condition "$\mathrm{supp}\,\varphi\subset\Omega$" is a non trivial one, since $\mathrm{supp}\,\varphi$ is closed and Ω is open. Functions satisfying this condition vanish at the boundary of Ω together with all their derivatives.

Given a sufficiently smooth function φ and a multi-index $\alpha\in\mathbb{N}^n$, we denote its partial derivatives by

$$D^\alpha\varphi = \frac{\partial^{\alpha_1+\ldots+\alpha_n}\varphi}{\partial x_1^{\alpha_1}\ldots\partial x_n^{\alpha_n}}.$$

Given two function $\varphi,\psi\in C_0^\infty(\Omega)$, the Gauß theorem for integrals yields for every multi-index $\alpha\in\mathbb{N}^n$ the identity

$$\int_\Omega D^\alpha\varphi\psi\mathrm{d}x = (-1)^{\alpha_1+\ldots+\alpha_n}\int_\Omega \varphi D^\alpha\psi\mathrm{d}x.$$

This identity motivates the definition of the weak derivatives:

Given two integrable function $\varphi, \psi \in L^1(\Omega)$ and a multi-index $\alpha \in \mathbb{N}^n$, ψ is called the α-th *weak derivative* of φ if and only if the identity

$$\int_\Omega \psi \rho \mathrm{d}x = (-1)^{\alpha_1+\ldots+\alpha_n} \int_\Omega \varphi D^\alpha \rho \mathrm{d}x$$

holds for all functions $\rho \in C_0^\infty(\Omega)$. In this case we write

$$\psi = D^\alpha \varphi.$$

REMARK I.2.3. For smooth functions, the notions of classical and weak derivatives coincide. Yet, there are functions which are not differentiable in the classical sense but which have a weak derivative (cf. Example I.2.4 below).

EXAMPLE I.2.4. The function $|x|$ is not differentiable in $(-1, 1)$, but it is differentiable in the weak sense. Its weak derivative is the piecewise constant function which equals -1 on $(-1, 0)$ and 1 on $(0, 1)$.

I.2.5. Sobolev spaces and norms. We will frequently use the following *Sobolev spaces* and norms:

$$H^k(\Omega) = \{\varphi \in L^2(\Omega) : D^\alpha \varphi \in L^2(\Omega) \text{ for all } \alpha \in \mathbb{N}^n \text{ with } \alpha_1 + \ldots + \alpha_n \le k\},$$

$$|\varphi|_k = \left\{ \sum_{\substack{\alpha \in \mathbb{N}^n \\ \alpha_1+\ldots+\alpha_n = k}} \|D^\alpha \varphi\|^2_{L^2(\Omega)} \right\}^{1/2},$$

$$\|\varphi\|_k = \left\{ \sum_{l=0}^{k} |\varphi|_l^2 \right\}^{1/2} = \left\{ \sum_{\substack{\alpha \in \mathbb{N}^n \\ \alpha_1+\ldots+\alpha_n \le k}} \|D^\alpha \varphi\|^2_{L^2(\Omega)} \right\}^{1/2},$$

$$H_0^1(\Omega) = \{\varphi \in H^1(\Omega) : \varphi = 0 \text{ on } \Gamma\},$$

$$L_0^2(\Omega) = \left\{ p \in L^2(\Omega) : \int_\Omega p = 0 \right\},$$

$$H^{\frac{1}{2}}(\Gamma) = \left\{ \psi \in L^2(\Gamma) : \psi = \varphi\Big|_\Gamma \text{ for some } \varphi \in H^1(\Omega) \right\},$$

$$\|\psi\|_{\frac{1}{2},\Gamma} = \inf \left\{ \|\varphi\|_1 : \varphi \in H^1(\Omega), \varphi\Big|_\Gamma = \psi \right\}.$$

Note that all derivatives are to be understood in the weak sense.

REMARK I.2.5. The space $H^{\frac{1}{2}}(\Gamma)$ is called *trace space* of $H^1(\Omega)$, its elements are called *traces* of functions in $H^1(\Omega)$. Except in one dimension, $n = 1$, H^1 functions are in general not continuous and do not admit point values (cf. Example I.2.6 below). A function, however, which is piecewise differentiable is in $H^1(\Omega)$ if and only if it is globally continuous. This is crucial for finite element functions.

EXAMPLE I.2.6. The function $|x|$ is not differentiable, but it is in $H^1((-1,1))$. In two dimension, the function $\ln(\ln(\sqrt{x_1^2+x_2^2}))$ is an example of an H^1-function that is not continuous and which does not admit a point value in the origin. In three dimensions, a similar example is given by $\ln(\sqrt{x_1^2+x_2^2+x_3^2})$.

EXAMPLE I.2.7. Consider the open unit ball

$$\Omega = \{x \in \mathbb{R}^n : x_1^2 + \ldots + x_n^2 < 1\}$$

in $\mathbb{R}^n$ and the functions

$$\varphi_\alpha(x) = \{x_1^2 + \ldots + x_n^2\}^{\frac{\alpha}{2}} \quad \alpha \in \mathbb{R}.$$

Then we have

$$\varphi_\alpha \in H^1(\Omega) \iff \begin{cases} \alpha \geq 0 & \text{if } n = 2, \\ \alpha > 1 - \frac{n}{2} & \text{if } n > 2. \end{cases}$$

I.2.6. Friedrichs and Poincaré inequalities. The following inequalities are fundamental:

$$\begin{aligned} &\|\varphi\|_0 \leq c_\Omega |\varphi|_1 && \text{for all } \varphi \in H_0^1(\Omega), \\ &&& \textit{Friedrichs inequality} \\ &\|\varphi\|_0 \leq c'_\Omega |\varphi|_1 && \text{for all } \varphi \in H^1(\Omega) \cap L_0^2(\Omega) \\ &&& \textit{Poincaré inequality.} \end{aligned}$$

The constants c_Ω and c'_Ω depend on the domain Ω and are proportional to its diameter.

I.2.7. Finite element partitions. The finite element discretizations are based on partitions of the domain Ω into non-overlapping simple sub-domains. The collection of these sub-domains is called a *partition* and is labeled $\mathcal{T}$. The members of $\mathcal{T}$, i.e. the sub-domains, are called *elements* and are labeled K.

Any partition $\mathcal{T}$ has to satisfy the following conditions:

- $\Omega \cup \Gamma$ is the union of all elements in $\mathcal{T}$.

- *(Affine equivalence)* Each $K \in \mathcal{T}$ is either a triangle or a parallelogram, if $n = 2$, or a tetrahedron or a parallelepiped, if $n = 3$.
- *(Admissibility)* Any two elements in $\mathcal{T}$ are either disjoint or share a vertex or a complete edge or – if $n = 3$ – a complete face.
- *(Shape-regularity)* For any element K, the ratio of its diameter h_K to the diameter ρ_K of the largest ball inscribed into K is bounded independently of K.

REMARK I.2.8. In two dimensions, $n = 2$, shape regularity means that the smallest angles of all elements stay bounded away from zero. In practice one usually not only considers a single partition $\mathcal{T}$, but complete families of partitions which are often obtained by successive local or global refinements. Then, the ratio h_K/ρ_K must be bounded *uniformly* with respect to *all elements and all partitions*.

For any partition $\mathcal{T}$ we denote by $h_\mathcal{T}$ or simply h the maximum of the element diameters:

$$h = h_\mathcal{T} = \max\{h_K : K \in \mathcal{T}\}.$$

REMARK I.2.9. In two dimensions triangles and parallelograms may be mixed (cf. Figure I.2.1). In three dimensions tetrahedrons and parallelepipeds can be mixed provided prismatic elements are also incorporated. The condition of affine equivalence may be dropped. It, however, considerably simplifies the analysis since it implies constant Jacobians for all element transformations.

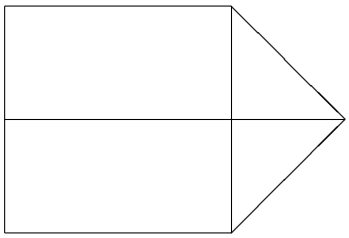

FIGURE I.2.1. Mixture of triangular and quadrilateral elements

I.2.8. Finite element spaces. For any multi-index $\alpha \in \mathbb{N}^n$ we set for abbreviation

$$|\alpha|_1 = \alpha_1 + \ldots + \alpha_n,$$

$$|\alpha|_\infty = \max\{\alpha_i : 1 \le i \le n\},$$
$$x^\alpha = x_1^{\alpha_1} \cdot \ldots \cdot x_n^{\alpha_n}.$$

Denote by

$$\widehat{K} = \{\widehat{x} \in \mathbb{R}^d : x_1 + \ldots + x_n \le 1, x_i \ge 0, 1 \le i \le n\}$$

the *reference simplex* for a partition into triangles or tetrahedra and by

$$\widehat{K} = [0,1]^n$$

the *reference cube* for a partition into parallelograms or parallelepipeds. Then every element $K \in \mathcal{T}$ is the image of $\widehat{K}$ under an affine mapping F_K. For every integer number k set

$$R_k(\widehat{K}) = \begin{cases} \operatorname{span}\{x^\alpha : |\alpha|_1 \le k\} & \text{if } K \text{ is the reference simplex,} \\ \operatorname{span}\{x^\alpha : |\alpha|_\infty \le k\} & \text{if } K \text{ is the reference cube} \end{cases}$$

and set

$$R_k(K) = \left\{\widehat{p} \circ F_K^{-1} : \widehat{p} \in \widehat{R}_k\right\}.$$

With this notation we define finite element spaces by

$$\begin{aligned} S^{k,-1}(\mathcal{T}) &= \left\{\varphi : \Omega \to \mathbb{R} : \varphi\Big|_K \in R_k(K) \ \text{ for all } K \in \mathcal{T}\right\}, \\ S^{k,0}(\mathcal{T}) &= S^{k,-1}(\mathcal{T}) \cap C(\overline{\Omega}), \\ S_0^{k,0}(\mathcal{T}) &= S^{k,0}(\mathcal{T}) \cap H_0^1(\Omega) = \{\varphi \in S^{k,0}(\mathcal{T}) : \varphi = 0 \ \text{ on } \Gamma\}. \end{aligned}$$

Note, that k may be 0 for the first space, but must be at least 1 for the second and third space.

Example I.2.10. For the reference triangle, we have

$$R_1(\widehat{K}) = \operatorname{span}\{1, x_1, x_2\},$$
$$R_2(\widehat{K}) = \operatorname{span}\{1, x_1, x_2, x_1^2, x_1x_2, x_2^2\}.$$

For the reference square on the other hand, we have

$$R_1(\widehat{K}) = \operatorname{span}\{1, x_1, x_2, x_1x_2\},$$
$$R_2(\widehat{K}) = \operatorname{span}\{1, x_1, x_2, x_1x_2, x_1^2, x_1^2x_2, x_1^2x_2^2, x_1x_2^2, x_2^2\}.$$

I.2.9. Approximation properties. The finite element spaces defined above satisfy the following approximation properties:

$$\begin{aligned} \inf_{\varphi_\mathcal{T} \in S^{k,-1}(\mathcal{T})} \|\varphi - \varphi_\mathcal{T}\|_0 &\le ch^{k+1}|\varphi|_{k+1} \quad \varphi \in H^{k+1}(\Omega),\ k \in \mathbb{N}, \\ \inf_{\varphi_\mathcal{T} \in S^{k,0}(\mathcal{T})} |\varphi - \varphi_\mathcal{T}|_j &\le ch^{k+1-j}|\varphi|_{k+1} \quad \varphi \in H^{k+1}(\Omega), \\ &\qquad\qquad\qquad\qquad\qquad\qquad j \in \{0,1\},\ k \in \mathbb{N}^*, \end{aligned}$$

$$\inf_{\varphi_{\mathcal{T}} \in S_0^{k,0}(\mathcal{T})} |\varphi - \varphi_{\mathcal{T}}|_j \le c h^{k+1-j} |\varphi|_{k+1} \quad \varphi \in H^{k+1}(\Omega) \cap H_0^1(\Omega),$$
$$j \in \{0,1\},\ k \in \mathbb{N}^*.$$

I.2.10. Nodal shape functions. $\mathcal{N}$ denotes the set of all element vertices.

For any vertex $x \in \mathcal{N}$ the associated *nodal shape function* is denoted by λ_x. It is the unique function in $S^{1,0}(\mathcal{T})$ that equals 1 at vertex x and that vanishes at all other vertices $y \in \mathcal{N} \backslash \{x\}$.

The support of a nodal shape function λ_x is denoted by ω_x and consists of all elements that share the vertex x (cf. Figure I.2.2).

FIGURE I.2.2. Some examples of domains ω_x

The nodal shape functions can easily be computed elementwise from the coordinates of the element's vertices.

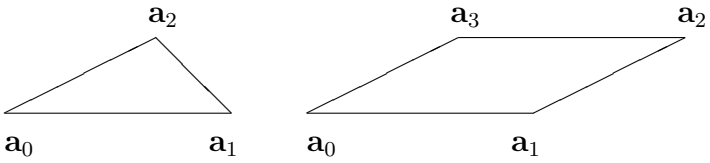

FIGURE I.2.3. Enumeration of vertices of triangles and parallelograms

EXAMPLE I.2.11. (1) Consider a triangle K with vertices $\mathbf{a}_0, \ldots, \mathbf{a}_2$ numbered counterclockwise (cf. Figure I.2.3). Then the restrictions to K of the nodal shape functions $\lambda_{\mathbf{a}_0}, \ldots, \lambda_{\mathbf{a}_2}$ are given by

$$\lambda_{\mathbf{a}_i}(x) = \frac{\det(x - \mathbf{a}_{i+1}\,,\, \mathbf{a}_{i+2} - \mathbf{a}_{i+1})}{\det(\mathbf{a}_i - \mathbf{a}_{i+1}\,,\, \mathbf{a}_{i+2} - \mathbf{a}_{i+1})} \quad i = 0, 1, 2,$$

where all indices have to be taken modulo 3.
(2) Consider a parallelogram K with vertices $\mathbf{a}_0, \ldots, \mathbf{a}_3$ numbered counterclockwise (cf. Figure I.2.3). Then the restrictions to K of the nodal shape functions $\lambda_{\mathbf{a}_0}, \ldots, \lambda_{\mathbf{a}_3}$ are given by

$$\lambda_{\mathbf{a}_i}(x) = \frac{\det(x - \mathbf{a}_{i+2}\,,\, \mathbf{a}_{i+3} - \mathbf{a}_{i+2})}{\det(\mathbf{a}_i - \mathbf{a}_{i+2}\,,\, \mathbf{a}_{i+3} - \mathbf{a}_{i+2})} \cdot \frac{\det(x - \mathbf{a}_{i+2}\,,\, \mathbf{a}_{i+1} - \mathbf{a}_{i+2})}{\det(\mathbf{a}_i - \mathbf{a}_{i+2}\,,\, \mathbf{a}_{i+1} - \mathbf{a}_{i+2})}$$

$$i = 0, \dots, 3,$$

where all indices have to be taken modulo 4.
(3) Consider a tetrahedron K with vertices $\mathbf{a}_0, \dots, \mathbf{a}_3$ enumerated as in Figure I.2.4. Then the restrictions to K of the nodal shape functions $\lambda_{\mathbf{a}_0}, \dots, \lambda_{\mathbf{a}_3}$ are given by

$$\lambda_{\mathbf{a}_i}(x) = \frac{\det(x - \mathbf{a}_{i+1}\,,\, \mathbf{a}_{i+2} - \mathbf{a}_{i+1}\,,\, \mathbf{a}_{i+3} - \mathbf{a}_{i+1})}{\det(\mathbf{a}_i - \mathbf{a}_{i+1}\,,\, \mathbf{a}_{i+2} - \mathbf{a}_{i+1}\,,\, \mathbf{a}_{i+3} - \mathbf{a}_{i+1})} \quad i = 0, \dots, 3,$$

where all indices have to be taken modulo 4.
(4) Consider a parallelepiped K with vertices $\mathbf{a}_0, \dots, \mathbf{a}_7$ enumerated as in Figure I.2.4. Then the restrictions to K of the nodal shape functions $\lambda_{\mathbf{a}_0}, \dots, \lambda_{\mathbf{a}_7}$ are given by

$$\begin{aligned}\lambda_{\mathbf{a}_i}(x) = {} & \frac{\det(x - \mathbf{a}_{i+1}\,,\, \mathbf{a}_{i+3} - \mathbf{a}_{i+1}\,,\, \mathbf{a}_{i+5} - \mathbf{a}_{i+1})}{\det(\mathbf{a}_i - \mathbf{a}_{i+1}\,,\, \mathbf{a}_{i+3} - \mathbf{a}_{i+1}\,,\, \mathbf{a}_{i+5} - \mathbf{a}_{i+1})} \cdot \\ & \frac{\det(x - \mathbf{a}_{i+2}\,,\, \mathbf{a}_{i+3} - \mathbf{a}_{i+2}\,,\, \mathbf{a}_{i+6} - \mathbf{a}_{i+2})}{\det(\mathbf{a}_i - \mathbf{a}_{i+2}\,,\, \mathbf{a}_{i+3} - \mathbf{a}_{i+2}\,,\, \mathbf{a}_{i+6} - \mathbf{a}_{i+2})} \cdot \\ & \frac{\det(x - \mathbf{a}_{i+4}\,,\, \mathbf{a}_{i+5} - \mathbf{a}_{i+4}\,,\, \mathbf{a}_{i+6} - \mathbf{a}_{i+4})}{\det(\mathbf{a}_i - \mathbf{a}_{i+4}\,,\, \mathbf{a}_{i+5} - \mathbf{a}_{i+4}\,,\, \mathbf{a}_{i+6} - \mathbf{a}_{i+4})} \\ & i = 0, \dots, 7,\end{aligned}$$

where all indices have to be taken modulo 8.

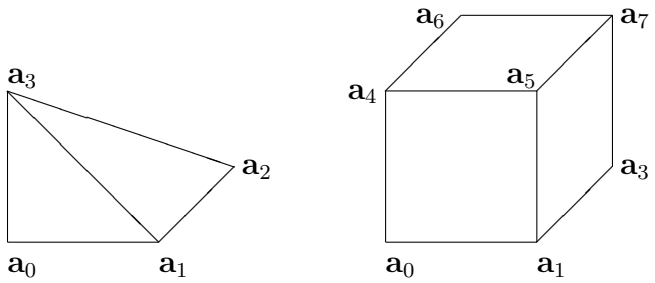

FIGURE I.2.4. Enumeration of vertices of tetrahedra and parallelepipeds (The vertex $\mathbf{a}_2$ of the parallelepiped is hidden.)

REMARK I.2.12. For every element (triangle, parallelogram, tetrahedron, or parallelepiped) the sum of all nodal shape functions corresponding to the element's vertices is identical equal to 1 on the element.

The functions λ_x, $x \in \mathcal{N}$, form a bases of $S^{1,0}(\mathcal{T})$. The bases of higher-order spaces $S^{k,0}(\mathcal{T})$, $k \geq 2$, consist of suitable products of functions λ_x corresponding to appropriate vertices x.

EXAMPLE I.2.13. (1) Consider a again a triangle K with its vertices numbered as in example I.2.11 (1). Then the nodal basis of $S^{2,0}(\mathcal{T})\Big|_K$ consists of the functions

$$\begin{aligned} \lambda_{\mathbf{a}_i}[\lambda_{\mathbf{a}_i} - \lambda_{\mathbf{a}_{i+1}} - \lambda_{\mathbf{a}_{i+2}}] \quad & i = 0, 1, 2 \\ 4\lambda_{\mathbf{a}_i}\lambda_{\mathbf{a}_{i+1}} \quad & i = 0, 1, 2, \end{aligned}$$

where the functions $\lambda_{\mathbf{a}_\ell}$ are as in example I.2.11 (1) and where all indices have to be taken modulo 3. An other basis of $S^{2,0}(\mathcal{T})\Big|_K$, called *hierarchical basis*, consists of the functions

$$\begin{aligned} \lambda_{\mathbf{a}_i} \quad & i = 0, 1, 2 \\ 4\lambda_{\mathbf{a}_i}\lambda_{\mathbf{a}_{i+1}} \quad & i = 0, 1, 2. \end{aligned}$$

(2) Consider a again a parallelogram K with its vertices numbered as in example I.2.11 (2). Then the nodal basis of $S^{2,0}(\mathcal{T})\Big|_K$ consists of the functions

$$\begin{aligned} \lambda_{\mathbf{a}_i}[\lambda_{\mathbf{a}_i} - \lambda_{\mathbf{a}_{i+1}} + \lambda_{\mathbf{a}_{i+2}} - \lambda_{\mathbf{a}_{i+3}}] \quad & i = 0, \dots, 3 \\ 4\lambda_{\mathbf{a}_i}[\lambda_{\mathbf{a}_{i+1}} - \lambda_{\mathbf{a}_{i+2}}] \quad & i = 0, \dots, 3 \\ 16\lambda_{\mathbf{a}_0}\lambda_{\mathbf{a}_2} & \end{aligned}$$

where the functions $\lambda_{\mathbf{a}_\ell}$ are as in example I.2.11 (2) and where all indices have to be taken modulo 4. The *hierarchical basis* of $S^{2,0}(\mathcal{T})\Big|_K$ consists of the functions

$$\begin{aligned} \lambda_{\mathbf{a}_i} \quad & i = 0, \dots, 3 \\ 4\lambda_{\mathbf{a}_i}[\lambda_{\mathbf{a}_{i+1}} - \lambda_{\mathbf{a}_{i+2}}] \quad & i = 0, \dots, 3 \\ 16\lambda_{\mathbf{a}_0}\lambda_{\mathbf{a}_2} \quad & . \end{aligned}$$

(3) Consider a again a tetrahedron K with its vertices numbered as in example I.2.11 (3). Then the nodal basis of $S^{2,0}(\mathcal{T})\Big|_K$ consists of the functions

$$\begin{aligned} \lambda_{\mathbf{a}_i}[\lambda_{\mathbf{a}_i} - \lambda_{\mathbf{a}_{i+1}} - \lambda_{\mathbf{a}_{i+2}} - \lambda_{\mathbf{a}_{i+3}}] \quad & i = 0, \dots, 3 \\ 4\lambda_{\mathbf{a}_i}\lambda_{\mathbf{a}_j} \quad & 0 \le i < j \le 3, \end{aligned}$$

where the functions $\lambda_{\mathbf{a}_\ell}$ are as in example I.2.11 (3) and where all indices have to be taken modulo 4. The hierarchical basis consists of the functions

$$\begin{aligned} \lambda_{\mathbf{a}_i} \quad & i = 0, \dots, 3 \\ 4\lambda_{\mathbf{a}_i}\lambda_{\mathbf{a}_j} \quad & 0 \le i < j \le 3. \end{aligned}$$

I.2.11. A quasi-interpolation operator. We will frequently use the *quasi-interpolation operator* $R_{\mathcal{T}} : L^1(\Omega) \to S_D^{1,0}(\mathcal{T})$ which is defined by

$$R_{\mathcal{T}}\varphi = \sum_{x \in \mathcal{N}_\Omega \cup \mathcal{N}_{\Gamma_N}} \lambda_x \frac{1}{|\omega_x|} \int_{\omega_x} \varphi \mathrm{d}x.$$

Here, $|\omega_x|$ denotes the area, if $n = 2$, respectively volume, if $n = 3$, of the set ω_x. The operator $R_{\mathcal{T}}$ has the following local approximation properties

$$\begin{aligned} \|\varphi - R_{\mathcal{T}}\varphi\|_{L^2(K)} &\leq c_1 h_K \|\varphi\|_{H^1(\widetilde{\omega}_K)}, \\ \|\varphi - R_{\mathcal{T}}\varphi\|_{L^2(\partial K)} &\leq c_2 h_K^{1/2} \|\varphi\|_{H^1(\widetilde{\omega}_K)}. \end{aligned}$$

Here, $\widetilde{\omega}_K$ denotes the set of all elements that share at least a vertex with K (cf. Figure I.2.5).

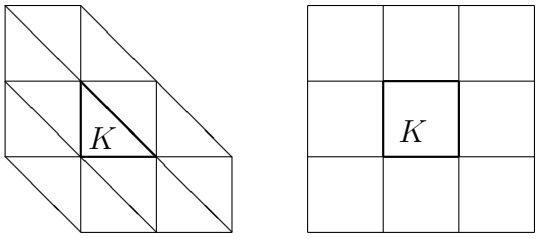

FIGURE I.2.5. Examples of domains $\widetilde{\omega}_K$

REMARK I.2.14. The operator $R_{\mathcal{T}}$ is called a quasi-interpolation operator since it *does not interpolate* a given function φ at the vertices $x \in \mathcal{N}$. In fact, point values *are not defined* for H^1-functions. For functions with more regularity which are at least in $H^2(\Omega)$, the situation is different. For those functions point values do exist and the classical nodal interpolation operator $I_{\mathcal{T}} : H^2(\Omega) \cap H_0^1(\Omega) \to S_0^{1,0}(\mathcal{T})$ can be defined by the relation $(I_{\mathcal{T}}(\varphi))(x) = \varphi(x)$ for all vertices $x \in \mathcal{N}$.

I.2.12. Bubble functions. For any element $K \in \mathcal{T}$ we define an *element bubble function* by

$$\psi_K = \alpha_K \prod_{x \in \mathcal{N}_K} \lambda_x \,,$$

$$\alpha_K = \begin{cases} 27 & \text{if } K \text{ is a triangle,} \\ 256 & \text{if } K \text{ is a tetrahedron,} \\ 16 & \text{if } K \text{ is a parallelogram,} \\ 64 & \text{if } K \text{ is a parallelepiped,} \end{cases}$$

where $\mathcal{N}_K$ is the set of all vertices of K. It has the following properties:

$$\begin{aligned} 0 \le \psi_K(x) \le 1 \qquad & \text{for all } x \in K, \\ \psi_K(x) = 0 \qquad & \text{for all } x \notin K, \\ \max_{x \in K} \psi_K(x) = 1. & \end{aligned}$$

We denote by $\mathcal{E}$ the set of all edges, if $n = 2$, and of all faces, if $n = 3$, of all elements in $\mathcal{T}$. With each edge respectively face $E \in \mathcal{E}$ we associate an *edge* respectively *face bubble function* by

$$\psi_E = \beta_E \prod_{x \in \mathcal{N}_E} \lambda_x \ ,$$

$$\beta_E = \begin{cases} 4 & \text{if } E \text{ is a line segment,} \\ 27 & \text{if } E \text{ is a triangle,} \\ 16 & \text{if } E \text{ is a parallelogram,} \end{cases}$$

where $\mathcal{N}_E$ is the set of all vertices of E. It has the following properties:

$$\begin{aligned} 0 \le \psi_E(x) \le 1 \qquad & \text{for all } x \in \omega_E, \\ \psi_E(x) = 0 \qquad & \text{for all } x \notin \omega_E, \\ \max_{x \in \omega_E} \psi_E(x) = 1. & \end{aligned}$$

Here ω_E is the union of all elements that share E (cf. Figure I.2.6). Note that ω_E consists of two elements, if E is not contained in the boundary Γ, and of exactly one element, if E is a subset of Γ.

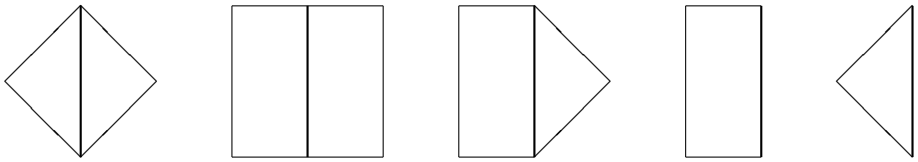

FIGURE I.2.6. Examples of domains ω_E (E is marked bold.)

With each edge respectively face $E \in \mathcal{E}$ we finally associate a unit vector $\mathbf{n}_E$ orthogonal to E and denote by $\mathbb{J}_E(\cdot)$ the jump across E in direction $\mathbf{n}_E$. If E is contained in the boundary Γ the orientation of $\mathbf{n}_E$ is fixed to be the one of the exterior normal. Otherwise it is not fixed.

REMARK I.2.15. $\mathbb{J}_E(\cdot)$ depends on the orientation of $\mathbf{n}_E$ but quantities of the form $\mathbb{J}_E(\mathbf{n}_E \cdot \varphi)$ are independent of this orientation.

CHAPTER II

Stationary linear problems

II.1. Discretization of the Stokes equations. A first attempt

II.1.1. The Poisson equation revisited. We recall the Poisson equation

$$\begin{aligned} -\Delta\varphi &= f \quad \text{in } \Omega \\ \varphi &= 0 \quad \text{on } \Gamma \end{aligned}$$

and its variational formulation: Find $\varphi \in H^1_0(\Omega)$ such that

$$\int_\Omega \nabla\varphi\nabla\psi \mathrm{d}x = \int_\Omega f\psi \mathrm{d}x$$

holds for all $\psi \in H^1_0(\Omega)$.

The Poisson equation and this variational formulation are equivalent in the sense that any solution of the Poisson equation is a solution of the variational problem and that conversely any solution of the variational problem which is twice continuously differentiable is a solution of the Poisson equation. Moreover, one can prove that the variational problem admits a unique solution.

Standard finite element discretizations simply choose a partition $\mathcal{T}$ of Ω and an integer $k \geq 1$ and replace in the variational problem $H^1_0(\Omega)$ by the finite dimensional space $S^{k,0}_0(\mathcal{T})$. This gives rise to a linear system of equations with a symmetric, positive definite, square matrix, called *stiffness matrix*. If, e.g. $k = 1$, the size of the resulting discrete problem is given by the number of vertices in $\mathcal{T}$ which do not lie on the boundary Γ.

II.1.2. A variational formulation of the Stokes equations. Now we look at the Stokes equations of §I.1.14 (p. 16) with no-slip boundary condition

$$\begin{aligned} -\Delta\mathbf{u} + \operatorname{grad} p &= \mathbf{f} \quad \text{in } \Omega \\ \operatorname{div}\mathbf{u} &= 0 \quad \text{in } \Omega \\ \mathbf{u} &= 0 \quad \text{on } \Gamma. \end{aligned}$$

The space

$$V = \{\mathbf{u} \in H^1_0(\Omega)^n : \operatorname{div}\mathbf{u} = 0\}$$

is a sub-space of $H^1_0(\Omega)^n$ which has similar properties. For every $p \in H^1(\Omega)$ and any $\mathbf{u} \in V$ the integration by parts formulae of §I.2.3 (p. 18)

imply that

$$\int_\Omega \nabla p \cdot \mathbf{u} \mathrm{d}x = -\int_\Omega p \operatorname{div} \mathbf{u} \mathrm{d}x = 0.$$

We therefore obtain the following variational formulation of the Stokes equations: Find $\mathbf{u} \in V$ such that

$$\int_\Omega \nabla \mathbf{u} : \nabla \mathbf{u} \mathrm{d}x = \int_\Omega \mathbf{f} \cdot \mathbf{v} \mathrm{d}x$$

holds for all $\mathbf{v} \in V$.

As for the Poisson problem, any velocity field which solves the Stokes equations also solves this variational problem. Moreover, one can prove that the variational problem admits a unique solution. Yet, we have lost the pressure! This is an apparent gap between differential equation and variational problem which is not present for the Poisson equation.

II.1.3. A naive discretization of the Stokes equations. Although the variational problem of the previous section does not incorporate the pressure, it nevertheless gives information about the velocity field. Therefore one may be attempted to use it as a starting point for a discretization process similar to the Poisson equation. This would yield a symmetric, positive definite system of linear equations for a discrete velocity field. This would have the appealing side-effect that standard solvers such as preconditioned conjugate gradient algorithms were available. The following example shows that this approach leads to a dead-end road.

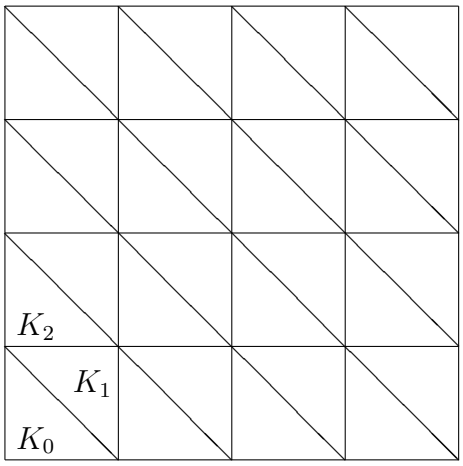

FIGURE II.1.1. Courant triangulation

EXAMPLE II.1.1. We consider the unit square $\Omega = (0,1)^2$ and divide it into N^2 squares of equal size. Each square is further divided into two triangles by connecting its top-left corner with its bottom-right corner

(cf. Figure II.1.1). This partition is known as *Courant triangulation*. The length of the triangles' short sides is $h = \frac{1}{N-1}$. For the discretization of the Stokes equations we replace V by $V(\mathcal{T}) = S_0^{1,0}(\mathcal{T})^2 \cap V$, i.e. we use continuous, piecewise linear, solenoidal finite element functions. Choose an arbitrary function $\mathbf{v}_\mathcal{T} \in V(\mathcal{T})$. We first consider the triangle K_0 which has the origin as a vertex. Since all its vertices are situated on the boundary, we have $\mathbf{v}_\mathcal{T} = 0$ on K_0. Next we consider the triangle K_1 which shares its longest side with K_0. Denote by z_1 the vertex of K_1 which lies in the interior of Ω. Taking into account that $\mathbf{v}_\mathcal{T} = 0$ on Γ and $\operatorname{div} \mathbf{v}_\mathcal{T} = 0$ in K_1, applying the integration by parts formulae of §I.2.3 (p. 18) and evaluating line integrals with the help of the trapezoidal rule, we conclude that

$$\begin{aligned} 0 &= \int_{K_1} \operatorname{div} \mathbf{v}_\mathcal{T} \mathrm{d}x \\ &= \int_{\partial K_1} \mathbf{v}_\mathcal{T} \cdot \mathbf{n}_{K_1} \mathrm{d}S \\ &= \frac{h}{2} \mathbf{v}_\mathcal{T}(z_1) \cdot \{\mathbf{e}_1 + \mathbf{e}_2\}. \end{aligned}$$

Here $\mathbf{n}_{K_1}$ is the unit exterior normal to K_1 and $\mathbf{e}_1$, $\mathbf{e}_2$ denote the canonical bases vectors in $\mathbb{R}^2$. Next we consider the triangle K_2, which is adjacent to the vertical boundary of Ω and to K_1. With the same arguments as for K_1 we conclude that

$$\begin{aligned} 0 &= \int_{K_2} \operatorname{div} \mathbf{v}_\mathcal{T} \mathrm{d}x \\ &= \int_{\partial K_2} \mathbf{v}_\mathcal{T} \cdot \mathbf{n}_{K_2} \mathrm{d}S \\ &= \underbrace{\frac{h}{2} \mathbf{v}_\mathcal{T}(z_1) \cdot \{\mathbf{e}_1 + \mathbf{e}_2\}}_{=0} - \frac{h}{2} \mathbf{v}_\mathcal{T}(z_1) \cdot \mathbf{e}_2 \\ &= -\frac{h}{2} \mathbf{v}_\mathcal{T}(z_1) \cdot \mathbf{e}_2. \end{aligned}$$

Since $\mathbf{v}_\mathcal{T}(z_1) \cdot \mathbf{e}_1 = -\mathbf{v}_\mathcal{T}(z_1) \cdot \mathbf{e}_2$ we obtain

$$\mathbf{v}_\mathcal{T}(z_1) = 0$$

and consequently $\mathbf{v}_\mathcal{T} = 0$ on $K_0 \cup K_1 \cup K_2$. Repeating this argument, we first conclude that $\mathbf{v}_\mathcal{T}$ vanishes on all squares that are adjacent to the left boundary of Ω and then that $\mathbf{v}_\mathcal{T}$ vanishes on all of Ω. Hence we have $V(\mathcal{T}) = \{0\}$. Thus this space is not suited for the approximation of V.

Remark II.1.2. One can prove that $S_0^{k,0}(\mathcal{T})^2 \cap V$ is suited for the approximation of V only if $k \geq 5$. This is no practical choice.

II.1.4. Possible remedies. The previous sections show that the Poisson and Stokes equations are fundamentally different and that the discretization of the Stokes problem is a much more delicate task than for the Poisson equation.

There are several possible remedies for the difficulties presented above:

- Relax the divergence constraint: This leads to mixed finite elements.
- Add consistent penalty terms: This leads to stabilized Petrov-Galerkin formulations.
- Relax the continuity condition: This leads to non-conforming methods.
- Look for an equivalent differential equation without the divergence constraint: This leads to the stream-function formulation.

In the following sections we will investigate all these possibilities and will see that each has its particular pitfalls.

II.2. Mixed finite element discretizations of the Stokes equations

II.2.1. Saddle-point formulation of the Stokes equations. We recall the Stokes equations with no-slip boundary condition

$$\begin{aligned} -\Delta \mathbf{u} + \operatorname{grad} p &= \mathbf{f} \quad \text{in } \Omega \\ \operatorname{div} \mathbf{u} &= 0 \quad \text{in } \Omega \\ \mathbf{u} &= 0 \quad \text{on } \Gamma. \end{aligned}$$

If p is any pressure solving the Stokes equations and if c is any constant, obviously $p + c$ is also an admissible pressure for the Stokes equations. Hence, the pressure is at most unique up to an additive constant. We try to fix this constant by the normalization

$$\int_\Omega p \mathrm{d}x = 0.$$

If $\mathbf{v} \in H_0^1(\Omega)^n$ and $p \in H^1(\Omega)$ are arbitrary, the integration by parts formulae of §I.2.3 (p. 18) imply that

$$\begin{aligned} \int_\Omega \nabla p \cdot \mathbf{v} \mathrm{d}x &= \int_\Gamma p \underbrace{\mathbf{v} \cdot \mathbf{n}}_{=0} \mathrm{d}S - \int_\Omega p \operatorname{div} \mathbf{v} \mathrm{d}x \\ &= -\int_\Omega p \operatorname{div} \mathbf{v} \mathrm{d}x. \end{aligned}$$

These observations lead to the following *mixed variational formulation of the Stokes equations*:

Find $\mathbf{u} \in H_0^1(\Omega)^n$ and $p \in L_0^2(\Omega)$ such that

$$\int_\Omega \nabla\mathbf{u} : \nabla\mathbf{v}\mathrm{d}x - \int_\Omega p \operatorname{div}\mathbf{v}\mathrm{d}x = \int_\Omega \mathbf{f}\cdot\mathbf{v}\mathrm{d}x \quad \text{for all } \mathbf{v} \in H_0^1(\Omega)^n$$
$$\int_\Omega q \operatorname{div}\mathbf{u}\mathrm{d}x = 0 \qquad \text{for all } q \in L_0^2(\Omega).$$

It has the following properties:

- Any solution $\mathbf{u}$, p of the Stokes equations is a solution of the above variational problem.
- Any solution $\mathbf{u}$, p of the variational problem which is sufficiently smooth is a solution of the Stokes equations.
- The variational problem is the Euler-Lagrange equation corresponding to the constrained minimization problem

$$\text{Minimize } \frac{1}{2}\int_\Omega |\nabla\mathbf{u}|^2\mathrm{d}x - \int_\Omega \mathbf{f}\cdot\mathbf{u}\mathrm{d}x$$
$$\text{subject to } \int_\Omega q \operatorname{div}\mathbf{u}\mathrm{d}x = 0 \text{ for all } q \in L_0^2(\Omega).$$

  Hence it is a *saddle-point problem*, i.e. $\mathbf{u}$ is a minimizer, p is a maximizer and is the Lagrange multiplicator corresponding the constraint $\operatorname{div}\mathbf{u} = 0$.

A deep mathematical result is:

The mixed variational formulation of the Stokes equations admits a unique solution $\mathbf{u}$, p. This solution depends continuously on the force $\mathbf{f}$, i.e.

$$|\mathbf{u}|_1 + \|p\|_0 \le c_\Omega \|\mathbf{f}\|_0$$

with a constant c_Ω which only depends on the domain Ω.

II.2.2. General structure of mixed finite element discretizations of the Stokes equations. We choose a partition $\mathcal{T}$ of Ω and two associated finite element spaces $X(\mathcal{T}) \subset H_0^1(\Omega)^n$ for the velocity and $Y(\mathcal{T}) \subset L_0^2(\Omega)$ for the pressure. Then we replace in the mixed variational formulation of the Stokes equations the space $H_0^1(\Omega)^n$ by $X(\mathcal{T})$ and the space $L_0^2(\Omega)$ by $X(\mathcal{T})$.

This gives rise to the following general *mixed finite element discretization of the Stokes equations*:

Find $\mathbf{u}_\mathcal{T} \in X(\mathcal{T})$ and $p_\mathcal{T} \in Y(\mathcal{T})$ such that

$$\int_\Omega \nabla \mathbf{u}_\mathcal{T} : \nabla \mathbf{v}_\mathcal{T} \mathrm{d}x - \int_\Omega p_\mathcal{T} \operatorname{div} \mathbf{v}_\mathcal{T} \mathrm{d}x = \int_\Omega \mathbf{f} \cdot \mathbf{v}_\mathcal{T} \mathrm{d}x \quad \text{for all } \mathbf{v}_\mathcal{T} \in X(\mathcal{T})$$

$$\int_\Omega q_\mathcal{T} \operatorname{div} \mathbf{u}_\mathcal{T} \mathrm{d}x = 0 \quad \text{for all } q_\mathcal{T} \in Y(\mathcal{T}).$$

REMARK II.2.1. Obviously, any particular mixed finite element discretization of the Stokes equations is determined by specifying the partition $\mathcal{T}$ and the spaces $X(\mathcal{T})$ and $Y(\mathcal{T})$. The condition $X(\mathcal{T}) \subset H_0^1(\Omega)^n$ implies that the discrete velocities are *globally continuous* and vanish on the boundary. The condition $Y(\mathcal{T}) \subset L_0^2(\Omega)$ implies that the discrete pressures have vanishing mean value, i.e. $\int_\Omega p_\mathcal{T} \mathrm{d}x = 0$ for all $p_\mathcal{T} \in Y(\mathcal{T})$. The condition $\int_\Omega q_\mathcal{T} \operatorname{div} \mathbf{u}_\mathcal{T} \mathrm{d}x = 0$ for all $q_\mathcal{T} \in Y(\mathcal{T})$ in general *does not imply* $\operatorname{div} \mathbf{u}_\mathcal{T} = 0$.

II.2.3. A first attempt. We consider the unit square $\Omega = (0,1)^2$ and the Courant triangulation of Example II.1.1 (p. 30). We choose

$$X(\mathcal{T}) = S_0^{1,0}(\mathcal{T})^2$$
$$Y(\mathcal{T}) = S^{0,-1}(\mathcal{T}) \cap L_0^2(\mathcal{T})$$

i.e. a continuous, piecewise linear velocity approximation and a piecewise constant pressure approximation on a triangular grid. Assume that $\mathbf{u}_\mathcal{T}$, $p_\mathcal{T}$ is a solution of the discrete problem. Then $\operatorname{div} \mathbf{u}_\mathcal{T}$ is piecewise constant. The condition $\int_\Omega q_\mathcal{T} \operatorname{div} \mathbf{u}_\mathcal{T} \mathrm{d}x = 0$ for all $q_\mathcal{T} \in Y(\mathcal{T})$ therefore implies $\operatorname{div} \mathbf{u}_\mathcal{T} = 0$. Hence we conclude from Example II.1.1 that $\mathbf{u}_\mathcal{T} = 0$. Thus this mixed finite element discretization has the same defects as the one of Example II.1.1.

II.2.4. A necessary condition for a well-posed mixed discretization. We consider an arbitrary mixed finite element discretization of the Stokes equations with spaces $X(\mathcal{T})$ for the velocity and $Y(\mathcal{T})$ for the pressure. A minimal requirement for such a discretization obviously is its well-posedness, i.e. that it admits a unique solution. Hence the stiffness matrix must be invertible.

Denote by $n_\mathbf{u}$ the dimension of $X(\mathcal{T})$ and by n_p the one of $Y(\mathcal{T})$ and choose arbitrary bases for these spaces. Then the stiffness matrix has the form

$$\begin{pmatrix} A & B \\ B^T & 0 \end{pmatrix}$$

with a symmetric positive definite $n_\mathbf{u} \times n_\mathbf{u}$ matrix A and a rectangular $n_\mathbf{u} \times n_p$ matrix B. Since A is symmetric positive definite, the stiffness matrix is invertible if and only if the matrix B has rank n_p. Since the

rank of a rectangular $k \times m$ matrix is at most $\min\{k, m\}$, this leads to the following necessary condition for a well-posed mixed discretization:

$$\boxed{n_{\mathbf{u}} \geq n_p.}$$

Let's have a look at the example of the previous section in the light of this condition. Denote by N^2 the number of the squares. An easy calculation then yields $n_{\mathbf{u}} = 2(N-1)^2$ and $n_p = 2N^2 - 1$. Hence we have $n_p > n_{\mathbf{u}}$ and the discretization cannot be well-posed.

II.2.5. A second attempt. As in §II.2.3 (p. 34) we consider the unit square and divide it into N^2 squares of equal size with N even. Contrary to §II.2.3 the squares are not further sub-divided. We choose the spaces

$$X(\mathcal{T}) = S_0^{1,0}(\mathcal{T})^2$$
$$Y(\mathcal{T}) = S^{0,-1}(\mathcal{T}) \cap L_0^2(\mathcal{T})$$

i.e. a continuous, piecewise bilinear velocity approximation and a piecewise constant pressure approximation on a quadrilateral grid. This discretization is often refered to as *Q1/Q0 element*.

A simple calculation yields the dimensions $n_{\mathbf{u}} = 2(N-1)^2$ and $n_p = N^2 - 1$. Hence the condition $n_{\mathbf{u}} \geq n_p$ of the previous section is satisfied provided $N \geq 4$.

-1	$+1$	-1	$+1$
$+1$	-1	$+1$	-1
-1	$+1$	-1	$+1$
$+1$	-1	$+1$	-1

FIGURE II.2.1. Checkerboard mode

Denote by $K_{i,j}$, $0 \leq i, j \leq N-1$, the square which has the lower left corner (ih, jh) where $h = \frac{1}{N-1}$. We now look at the particular pressure $\widehat{p}_{\mathcal{T}} \in Y(\mathcal{T})$ which has the constant value $(-1)^{i+j}$ on $K_{i,j}$ (cf. Figure II.2.1). It is frequently called *checkerboard mode*. Consider an arbitrary velocity $\mathbf{v}_{\mathcal{T}} \in X(\mathcal{T})$. Using the integration by parts formulae

of §I.2.3 (p. 18) and evaluating line integrals with the trapezoidal rule we obtain

$$\begin{aligned}
&\int_{K_{ij}} \widehat{p}_{\mathcal{T}} \operatorname{div} \mathbf{v}_{\mathcal{T}} \mathrm{d}x \\
&= (-1)^{i+j} \int_{K_{ij}} \operatorname{div} \mathbf{v}_{\mathcal{T}} dx \\
&= (-1)^{i+j} \int_{\partial K_{ij}} \mathbf{v}_{\mathcal{T}} \cdot \mathbf{n}_{K_{ij}} \mathrm{d}S \\
&= (-1)^{i+j} \frac{h}{2} \Big\{ (\mathbf{v}_{\mathcal{T}} \cdot \mathbf{e}_1)((i+1)h, jh) + (\mathbf{v}_{\mathcal{T}} \cdot \mathbf{e}_1)((i+1)h, (j+1)h) \\
&\qquad - (\mathbf{v}_{\mathcal{T}} \cdot \mathbf{e}_1)(ih, (j+1)h) - (\mathbf{v}_{\mathcal{T}} \cdot \mathbf{e}_1)(ih, jh) \\
&\qquad + (\mathbf{v}_{\mathcal{T}} \cdot \mathbf{e}_2)((i+1)h, (j+1)h) + (\mathbf{v}_{\mathcal{T}} \cdot \mathbf{e}_2)(ih, (j+1)h) \\
&\qquad - (\mathbf{v}_{\mathcal{T}} \cdot \mathbf{e}_2)(ih, jh) - (\mathbf{v}_{\mathcal{T}} \cdot \mathbf{e}_2)((i+1)h, jh) \Big\}.
\end{aligned}$$

Due to the homogeneous boundary condition, summation with respect to all indices i, j yields

$$\int_{\Omega} \widehat{p}_{\mathcal{T}} \operatorname{div} \mathbf{v}_{\mathcal{T}} \mathrm{d}x = 0.$$

Since $\mathbf{v}_{\mathcal{T}}$ was arbitrary, the solution of the discrete problem cannot be unique: One may add $\widehat{p}_{\mathcal{T}}$ to any pressure solution and obtains a new one.

This phenomenon is known as *checkerboard instability*. It shows that the condition of the previous section is only a necessary one and does not imply the well-posedness of the discrete problem.

II.2.6. The inf-sup condition. In 1974 *Franco Brezzi* published the following *necessary and sufficient* condition for the well-posedness of a mixed finite element discretization of the Stokes equations:

$$\inf_{p_{\mathcal{T}} \in Y(\mathcal{T}) \setminus \{0\}} \sup_{\mathbf{u}_{\mathcal{T}} \in X(\mathcal{T}) \setminus \{0\}} \frac{\int_\Omega p_{\mathcal{T}} \operatorname{div} \mathbf{u}_{\mathcal{T}} \mathrm{d}x}{|\mathbf{u}_{\mathcal{T}}|_1 \|p_{\mathcal{T}}\|_0} \geq \beta > 0.$$

A pair $X(\mathcal{T})$, $Y(\mathcal{T})$ of finite element spaces satisfying this condition is called *stable*. The same notion refers to the corresponding discretization.

Remark II.2.2. In practice one usually not only considers a single partition $\mathcal{T}$, but complete families of partitions which are often obtained by successive local or global refinements. Usually they are labeled $\mathcal{T}_h$ where the index h indicates that the mesh-size gets smaller and smaller. In this situation, *the above condition of Franco Brezzi*

must be satisfied uniformly, i.e. the number β must be independent of the mesh-size.

REMARK II.2.3. The above condition is known under various names. A prosaic one is *inf-sup condition*. Another one, mainly used in western countries, is *Babuška-Brezzi condition*. A third one, which is popular in eastern Europe and Russia, is *Ladyzhenskaya-Babuška-Brezzi condition* in short *LBB-condition*. The additional names honour older results of Olga Ladyzhenskaya and Ivo Babuška which, in a certain sense, lay the ground for the result of Franco Brezzi.

In the following sections we give a catalogue of various stable elements. It incorporates the most popular elements, but it is not complete. We distinguish between discontinuous and continuous pressure approximations. The first variant sometimes gives a better approximation of the incompressibility condition; the second variant often leads to smaller discrete systems while retaining the accuracy of a comparable discontinuous pressure approximation.

II.2.7. A stable low-order element with discontinuous pressure approximation. This discretization departs from the unstable Q1/Q0-element. It is based on the observation that the instability of the Q1/Q0-element is due to the fact that the velocity space cannot balance pressure jumps across element boundaries. As a remedy the velocity space is enriched by edge respectively face bubble functions (cf. §I.2.12 (p. 27)) which give control on the pressure jumps.

The pair
$$X(\mathcal{T}) = S_0^{1,0}(\mathcal{T})^n \oplus \operatorname{span}\{\psi_E \mathbf{n}_E : E \in \mathcal{E}\}$$
$$Y(\mathcal{T}) = S^{0,-1}(\mathcal{T}) \cap L_0^2(\Omega)$$
is stable.

Since the bubble functions ψ_E are contained in $S^{n,0}(\mathcal{T})$ we obtain as a corollary:

The pair
$$X(\mathcal{T}) = S_0^{n,0}(\mathcal{T})^n$$
$$Y(\mathcal{T}) = S^{0,-1}(\mathcal{T}) \cap L_0^2(\Omega)$$
is stable.

II.2.8. Stable higher-order elements with discontinuous pressure approximation. For the following result we assume that

$\mathcal{T}$ exclusively consists of triangles or tetrahedrons. Moreover, we denote for every positive integer ℓ by

$$\mathbb{P}_\ell = \operatorname{span}\{x_1^{\alpha_1} \cdot \ldots \cdot x_n^{\alpha_n} : \alpha_1 + \ldots + \alpha_n = \ell\}$$

the space of homogeneous polynomials of degree ℓ. With these restrictions and notations the following result can be proven:

For every $m \geq n$ the pair
$$X(\mathcal{T}) = \left[S_0^{m,0}(\mathcal{T}) \oplus \operatorname{span}\{\psi_K \varphi : K \in \mathcal{T}, \varphi \in \mathbb{P}_{m-2}\}\right]^n$$
$$Y(\mathcal{T}) = S^{m-1,-1}(\mathcal{T}) \cap L_0^2(\Omega)$$
is stable.

Since the element-bubble functions ψ_K are polynomials of degree $n+1$ we obtain as a corollary:

For every $m \geq n$ the pair
$$X(\mathcal{T}) = S_0^{n+m-1,0}(\mathcal{T})^n$$
$$Y(\mathcal{T}) = S^{m-1,-1}(\mathcal{T}) \cap L_0^2(\Omega)$$
is stable.

II.2.9. Stable low-order elements with continuous pressure approximation. Given a partition $\mathcal{T}$ we denote by $\mathcal{T}_{\frac{1}{2}}$ the refined partition which is obtained by connecting in every element the midpoints of its edges. With this notation the following results were established in the early 1980s:

The *mini element*
$$X(\mathcal{T}) = \left[S_0^{1,0}(\mathcal{T}) \oplus \operatorname{span}\{\psi_K : K \in \mathcal{T}\}\right]^n$$
$$Y(\mathcal{T}) = S^{1,0}(\mathcal{T}) \cap L_0^2(\Omega)$$
is stable.

The *Hood-Taylor element*
$$X(\mathcal{T}) = S_0^{2,0}(\mathcal{T})^n$$
$$Y(\mathcal{T}) = S^{1,0}(\mathcal{T}) \cap L_0^2(\Omega)$$
is stable.

The *modified Hood-Taylor element*

$$X(\mathcal{T}) = S_0^{1,0}(\mathcal{T}_{\frac{1}{2}})^n$$
$$Y(\mathcal{T}) = S^{1,0}(\mathcal{T}) \cap L_0^2(\Omega)$$

is stable.

II.2.10. Stable higher-order elements with continuous pressure approximation. The stability of higher order elements with continuous pressure approximation was established only in the early 1990s:

For every $k \geq 3$ the *higher order Hood-Taylor element*

$$X(\mathcal{T}) = S_0^{k,0}(\mathcal{T})^n$$
$$Y(\mathcal{T}) = S^{k-1,0}(\mathcal{T}) \cap L_0^2(\Omega)$$

is stable.

II.2.11. A priori error estimates. We consider a partition $\mathcal{T}$ with mesh-size h and a stable pair $X(\mathcal{T})$, $Y(\mathcal{T})$ of corresponding finite element spaces for the discretization of the Stokes equations. We denote by $k_{\mathbf{u}}$ and k_p the maximal polynomial degrees k and m such that $S_0^{k,0}(\mathcal{T})^n \subset X(\mathcal{T})$ and either $S^{m,-1}(\mathcal{T}) \cap L_0^2(\Omega) \subset Y(\mathcal{T})$ when using a discontinuous pressure approximation or $S^{m,0}(\mathcal{T}) \cap L_0^2(\Omega) \subset Y(\mathcal{T})$ when using a continuous pressure approximation. Set

$$k = \min\{k_{\mathbf{u}} - 1, k_p\}.$$

Further we denote by $\mathbf{u}$, p the unique solution of the saddle-point formulation of the Stokes equations (cf. §II.2.1 (p. 32)) and by $\mathbf{u}_{\mathcal{T}}$, $p_{\mathcal{T}}$ the unique solution of the discrete problem under consideration.

With these notations the following a priori error estimates can be proven:

Assume that $\mathbf{u} \in H^{k+2}(\Omega)^n \cap H_0^1(\Omega)^n$ and $p \in H^{k+1}(\Omega) \cap L_0^2(\Omega)$, then

$$|\mathbf{u} - \mathbf{u}_{\mathcal{T}}|_1 + \|p - p_{\mathcal{T}}\|_0 \leq c_1 h^{k+1} \|\mathbf{f}\|_k.$$

If in addition Ω is convex, then

$$\|\mathbf{u} - \mathbf{u}_{\mathcal{T}}\|_0 \leq c_2 h^{k+2} \|\mathbf{f}\|_k.$$

The constants c_1 and c_2 only depend on the domain Ω.

EXAMPLE II.2.4. Table II.2.1 collects the numbers $k_{\mathbf{u}}$, k_p, and k for the examples of the previous sections.

TABLE II.2.1. Parameters of various stable elements

pair	$k_{\mathbf{u}}$	k_p	k
§II.2.7	1	0	0
§II.2.8	m	$m-1$	$m-1$
mini element	1	1	0
Hood-Taylor element	2	1	1
modified Hood-Taylor element	1	1	0
higher order Hood-Taylor element	k	$k-1$	$k-1$

REMARK II.2.5. The above regularity assumptions are not realistic. For a convex polygonal domain, one in general only has $\mathbf{u} \in H^2(\Omega)^n \cap H_0^1(\Omega)^n$ and $p \in H^1(\Omega) \cap L^2(\Omega)$. Therfore, independently of the actual discretization, one in general only gets an $O(h)$ error estimate. If Ω is not convex, but has re-entrant corners, the situation is even worse: One in general only gets an $O(h^\alpha)$ error estimate with an exponent $\frac{1}{2} \le \alpha < 1$ which depends on the largest interior angle at a boundary vertex of Ω. This poor convergence behaviour can only be remedied by an adaptive grid refinement based on an a posteriori error control.

REMARK II.2.6. The differentiation index of the pressure always is one less than the differentiation index of the velocity. Therefore, discretizations with $k_p = k_{\mathbf{u}} - 1$ are optimal in that they do not waste degrees of freedom by choosing too large a velocity space or too small a pressure space.

II.3. Petrov-Galerkin stabilization

II.3.1. Motivation. The results of §II.2 show that one can stabilize a given pair of finite element spaces by either reducing the number of degrees of freedom in the pressure space or by increasing the number of degrees of freedom in the velocity space. This result is reassuring but sometimes not practical since it either reduces the accuracy of the pressure or increases the number of unknowns and thus the computational work. Sometimes one would prefer to stick to a given pair of spaces and to have at hand a different stabilization process which does neither change the number of unknowns nor the accuracy of the finite element spaces. A particular example for this situation is the popular *equal order interpolation* where $X(\mathcal{T}) = S_0^{m,0}(\mathcal{T})^n$ and $Y(\mathcal{T}) = S^{m,0}(\mathcal{T}) \cap L_0^2(\Omega)$.

In this section we will devise a stabilization process with the desired properties. For its motivation we will have another look at the mini element of §II.2.9 (p. 38).

II.3.2. The mini element revisited. We consider a *triangulation* $\mathcal{T}$ of a *two dimensional* domain Ω and set for abbreviation

$$B(\mathcal{T}) = \big[\operatorname{span}\{\psi_K : K \in \mathcal{T}\}\big]^2.$$

The discretization of the Stokes equations with the mini element of §II.2.9 (p. 38) then takes the form:
Find $\mathbf{u}_\mathcal{T} \in S_0^{1,0}(\mathcal{T})^2 \oplus B(\mathcal{T})$ and $p_\mathcal{T} \in S^{1,0}(\mathcal{T}) \cap L_0^2(\Omega)$ such that

$$\int_\Omega \nabla \mathbf{u}_\mathcal{T} : \nabla \mathbf{v}_\mathcal{T} \mathrm{d}x - \int_\Omega p_\mathcal{T} \operatorname{div} \mathbf{v}_\mathcal{T} \mathrm{d}x = \int_\Omega \mathbf{f} \cdot \mathbf{v}_\mathcal{T} \mathrm{d}x \qquad \text{for all } \mathbf{v}_\mathcal{T} \in S_0^{1,0}(\mathcal{T})^2 \oplus B(\mathcal{T})$$

$$\int_\Omega q_\mathcal{T} \operatorname{div} \mathbf{u}_\mathcal{T} \mathrm{d}x = 0 \qquad \text{for all } q_\mathcal{T} \in S^{1,0}(\mathcal{T}) \cap L_0^2(\Omega).$$

For simplicity we assume that $\mathbf{f} \in L^2(\Omega)^2$.

We split the velocity $\mathbf{u}_\mathcal{T}$ in a "linear part" $\mathbf{u}_{\mathcal{T},L} \in S_0^{1,0}(\mathcal{T})^2$ and a "bubble part" $\mathbf{u}_{\mathcal{T},B} \in B(\mathcal{T})$:

$$\mathbf{u}_\mathcal{T} = \mathbf{u}_{\mathcal{T},L} + \mathbf{u}_{\mathcal{T},B}.$$

Since $\mathbf{u}_{\mathcal{T},B}$ vanishes on the element boundaries, integration by parts element-wise yields for every $\mathbf{v}_\mathcal{T} \in S_0^{1,0}(\mathcal{T})^2$

$$\begin{aligned}\int_\Omega \nabla \mathbf{u}_{\mathcal{T},B} : \nabla \mathbf{v}_\mathcal{T} \mathrm{d}x &= \sum_{K \in \mathcal{T}} \int_K \nabla \mathbf{u}_{\mathcal{T},B} : \nabla \mathbf{v}_\mathcal{T} \mathrm{d}x \\ &= -\sum_{K \in \mathcal{T}} \int_K \mathbf{u}_{\mathcal{T},B} \cdot \underbrace{\Delta \mathbf{v}_\mathcal{T}}_{=0} \mathrm{d}x \\ &= 0.\end{aligned}$$

Hence, we have

$$\int_\Omega \nabla \mathbf{u}_{\mathcal{T},L} : \nabla \mathbf{v}_\mathcal{T} \mathrm{d}x - \int_\Omega p_\mathcal{T} \operatorname{div} \mathbf{v}_\mathcal{T} \mathrm{d}x = \int_\Omega \mathbf{f} \cdot \mathbf{v}_\mathcal{T} \mathrm{d}x \qquad \text{for all } \mathbf{v}_\mathcal{T} \in S_0^{1,0}(\mathcal{T})^2.$$

The bubble part of the velocity has the representation

$$\mathbf{u}_{\mathcal{T},B} = \sum_{K \in \mathcal{T}} \boldsymbol{\alpha}_K \psi_K$$

with coefficients $\boldsymbol{\alpha}_K \in \mathbb{R}^2$. Inserting the test function $\mathbf{v}_\mathcal{T} = \mathbf{e}_i \psi_K$ with $i \in \{1, 2\}$ and $K \in \mathcal{T}$ in the momentum equation, we obtain

$$\begin{aligned}&\int_K \mathbf{f} \cdot \mathbf{e}_i \psi_K \mathrm{d}x \\ &\quad = \int_\Omega \mathbf{f} \cdot (\psi_K \mathbf{e}_i) \mathrm{d}x\end{aligned}$$

$$
\begin{aligned}
&= \underbrace{\int_\Omega \nabla \mathbf{u}_{\mathcal{T},L} : \nabla(\psi_K \mathbf{e}_i)\mathrm{d}x}_{=0} + \underbrace{\int_\Omega \nabla \mathbf{u}_{\mathcal{T},B} : \nabla(\psi_K \mathbf{e}_i)\mathrm{d}x}_{=\boldsymbol{\alpha}_{K,i}\int_K |\nabla\psi_K|^2\mathrm{d}x} \\
&\quad - \underbrace{\int_\Omega p_\mathcal{T} \operatorname{div}(\psi_K \mathbf{e}_i)\mathrm{d}x}_{=-\int_K \frac{\partial p_\mathcal{T}}{\partial x_i}\psi_K \mathrm{d}x} \\
&= \boldsymbol{\alpha}_{K,i}\int_K |\nabla\psi_K|^2\mathrm{d}x + \int_K \frac{\partial p_\mathcal{T}}{\partial x_i}\psi_K \mathrm{d}x, \quad i = 1,2,
\end{aligned}
$$

and thus

$$
\begin{aligned}
\boldsymbol{\alpha}_K = \left\{\int_K [\mathbf{f} - \nabla p_\mathcal{T}]\psi_K \mathrm{d}x\right\}\left\{\int_K |\nabla\psi_K|^2\mathrm{d}x\right\}^{-1} \\
\text{for all } K \in \mathcal{T}.
\end{aligned}
$$

For abbreviation we set

$$
\widetilde{\gamma}_K = \left\{\int_K |\nabla\psi_K|^2\mathrm{d}x\right\}^{-1}.
$$

Next we insert the representation of $\mathbf{u}_{\mathcal{T},B}$ in the continuity equation. This yields for all $q_\mathcal{T} \in S^{1,0}(\mathcal{T}) \cap L^2_0(\Omega)$

$$
\begin{aligned}
0 &= \int_\Omega q_\mathcal{T} \operatorname{div} \mathbf{u}_\mathcal{T}\mathrm{d}x \\
&= \int_\Omega q_\mathcal{T} \operatorname{div} \mathbf{u}_{\mathcal{T},L}\mathrm{d}x + \sum_{K\in\mathcal{T}} \underbrace{\int_K q_\mathcal{T} \operatorname{div}(\boldsymbol{\alpha}_K\psi_K)\mathrm{d}x}_{=-\int_K \psi_K \boldsymbol{\alpha}_K\cdot\nabla q_\mathcal{T}\mathrm{d}x} \\
&= \int_\Omega q_\mathcal{T} \operatorname{div} \mathbf{u}_{\mathcal{T},L}\mathrm{d}x \\
&\quad - \sum_{K\in\mathcal{T}} \widetilde{\gamma}_K \left\{\int_K \psi_K \nabla q_\mathcal{T}\mathrm{d}x\right\}\cdot\left\{\int_K \psi_K[\mathbf{f} - \nabla p_\mathcal{T}]\mathrm{d}x\right\}.
\end{aligned}
$$

This proves that $\mathbf{u}_{\mathcal{T},L} \in S^{1,0}_0(\mathcal{T})^2$, $p_\mathcal{T} \in S^{1,0}(\mathcal{T}) \cap L^2_0(\Omega)$ solve the modified problem

$$
\begin{aligned}
\int_\Omega \nabla \mathbf{u}_{\mathcal{T},L} : \nabla \mathbf{v}_\mathcal{T}\mathrm{d}x - \int_\Omega p_\mathcal{T} \operatorname{div} \mathbf{v}_\mathcal{T}\mathrm{d}x &= \int_\Omega \mathbf{f}\cdot\mathbf{v}_\mathcal{T}\mathrm{d}x \\
&\qquad \text{for all } \mathbf{v}_\mathcal{T} \in S^{1,0}_0(\mathcal{T})^2 \\
\int_\Omega q_\mathcal{T} \operatorname{div} \mathbf{u}_{\mathcal{T},L}\mathrm{d}x + c_\mathcal{T}(p_\mathcal{T}, q_\mathcal{T}) &= \chi_\mathcal{T}(q_\mathcal{T}) \\
&\qquad \text{for all } q_\mathcal{T} \in S^{1,0}(\mathcal{T}) \cap L^2_0(\Omega)
\end{aligned}
$$

where

$$c_{\mathcal{T}}(p_{\mathcal{T}}, q_{\mathcal{T}}) = \sum_{K\in\mathcal{T}} \gamma_K \int_K \nabla p_{\mathcal{T}} \cdot \nabla q_{\mathcal{T}} \mathrm{d}x$$
$$\chi_{\mathcal{T}}(q_{\mathcal{T}}) = \sum_{K\in\mathcal{T}} \widetilde{\gamma}_K \left\{\int_K \psi_K \nabla q_{\mathcal{T}} \mathrm{d}x\right\} \cdot \left\{\int_K \psi_K \mathbf{f} \mathrm{d}x\right\}$$
$$\gamma_K = \widetilde{\gamma}_K |K|^{-1} \left\{\int_K \psi_K \mathrm{d}x\right\}^2.$$

Note that γ_K is proportional to h_K^2.

The new problem can be interpreted as a P1/P1 discretization of the differential equation

$$\begin{aligned} -\Delta \mathbf{u} + \operatorname{grad} p &= \mathbf{f} && \text{in } \Omega \\ \operatorname{div} \mathbf{u} - \alpha \Delta p &= -\alpha \operatorname{div} \mathbf{f} && \text{in } \Omega \\ \mathbf{u} &= 0 && \text{on } \Gamma \end{aligned}$$

with a suitable *penalty parameter* α. Taking into account that $\Delta \mathbf{u}_{\mathcal{T}}$ vanishes elementwise, the discrete problem does not change if we also add the term $\alpha \Delta \operatorname{div} \mathbf{u}$ to the left-hand side of the second equation. This shows that in total we may add the divergence of the momentum equation as a penalty. Since the penalty vanishes for the exact solution of the Stokes equations, this approach is also called *consistent penalty*.

II.3.3. General form of Petrov-Galerkin stabilizations. Given a partition $\mathcal{T}$ of Ω we choose two corresponding finite element spaces $X(\mathcal{T}) \subset H_0^1(\Omega)^n$ and $Y(\mathcal{T}) \subset L_0^2(\Omega)$. For every element $K \in \mathcal{T}$ and every edge respectively face $E \in \mathcal{E}$ we further choose non-negative parameters δ_K and δ_E.

Recalling that $\mathbb{J}_E(\cdot)$ denotes the jump across E, we therefore consider the following discrete problem:

Find $\mathbf{u}_{\mathcal{T}} \in X(\mathcal{T})$ and $p_{\mathcal{T}} \in Y(\mathcal{T})$ such that

$$\int_\Omega \nabla \mathbf{u}_{\mathcal{T}} : \nabla \mathbf{v}_{\mathcal{T}} \mathrm{d}x - \int_\Omega p_{\mathcal{T}} \operatorname{div} \mathbf{v}_{\mathcal{T}} \mathrm{d}x = \int_\Omega \mathbf{f} \cdot \mathbf{v}_{\mathcal{T}} \mathrm{d}x$$
$$\int_\Omega q_{\mathcal{T}} \operatorname{div} \mathbf{u}_{\mathcal{T}} \mathrm{d}x$$
$$+ \sum_{K\in\mathcal{T}} \delta_K h_K^2 \int_K [-\Delta \mathbf{u}_{\mathcal{T}} + \nabla p_{\mathcal{T}}] \cdot \nabla q_{\mathcal{T}} \mathrm{d}x$$
$$+ \sum_{E\in\mathcal{E}} \delta_E h_E \int_E \mathbb{J}_E(p_{\mathcal{T}}) \mathbb{J}_E(q_{\mathcal{T}}) \mathrm{d}S = \sum_{K\in\mathcal{T}} \delta_K h_K^2 \int_K \mathbf{f} \cdot \nabla q_{\mathcal{T}} \mathrm{d}x$$

for all $\mathbf{v}_{\mathcal{T}} \in X(\mathcal{T})$ and all $q_{\mathcal{T}} \in Y(\mathcal{T})$.

REMARK II.3.1. The terms involving δ_K correspond to the equation

$$\operatorname{div}[-\Delta \mathbf{u} + \nabla p] = \operatorname{div} \mathbf{f}.$$

The terms involving δ_E are only relevant for discontinuous pressure approximations. When using a continuous pressure approximation they vanish.

II.3.4. Choice of stabilization parameters. With the notations of the previous section we set

$$\delta_{\max} = \max \left\{ \max_{K \in \mathcal{T}} \delta_K \, , \, \max_{E \in \mathcal{E}} \delta_E \right\},$$

$$\delta_{\min} = \begin{cases} \min \left\{ \min\limits_{K \in \mathcal{T}} \delta_K \, , \, \min\limits_{E \in \mathcal{E}} \delta_E \right\} & \text{if pressures are discontinuous,} \\ \min\limits_{K \in \mathcal{T}} \delta_K & \text{if pressures are continuous.} \end{cases}$$

A good choice of the stabilization parameters then is determined by the condition

$$\boxed{\delta_{\max} \approx \delta_{\min}.}$$

II.3.5. Choice of spaces. The a priori error estimates are optimal when the polynomial degree of the pressure approximation is one less than the polynomial degree of the velocity approximation. Hence a standard choice is

$$\boxed{\begin{aligned} X(\mathcal{T}) &= S_0^{k,0}(\mathcal{T}) \\ Y(\mathcal{T}) &= \begin{cases} S^{k-1,0}(\mathcal{T}) \cap L_0^2(\Omega) & \text{continuous pressure approximation} \\ S^{k-1,-1}(\mathcal{T}) \cap L_0^2(\Omega) & \text{discontinuous pressure approximation} \end{cases} \end{aligned}}$$

where $k \geq 1$ for discontinuous pressure approximations and $k \geq 2$ for continuous pressure approximations. These choices yield $\mathrm{O}(h^k)$ error estimates provided the solution of the Stokes equations is sufficiently smooth.

II.3.6. Structure of the discrete problem. The stiffness matrix of a Petrov-Galerkin scheme has the form

$$\begin{pmatrix} A & B \\ -B^T & C \end{pmatrix}$$

with symmetric, positive definite, square matrices A and C and a rectangular matrix B. The entries of C are by a factor of h^2 smaller than

those of A.
Recall that $C = 0$ for the discretizations of §II.2.

II.4. Non-conforming methods

II.4.1. Motivation. Up to now we have always imposed the condition $X(\mathcal{T}) \subset H^1_0(\Omega)^n$, i.e. the discrete velocities had to be continuous. Now, we will relax this condition. We hope that we will be rewarded by less difficulties with respect to the incompressibility constraint.

One can prove that, nevertheless, the velocities must enjoy some minimal continuity: On each edge the mean value of the velocity jump across this edge must vanish. Otherwise the consistency error will be at least of order O(1).

II.4.2. The Crouzeix-Raviart element. We consider a *triangulation* $\mathcal{T}$ of a *two dimensional* domain Ω and set

$$X(\mathcal{T}) = \Big\{ \mathbf{v}_\mathcal{T} : \Omega \to \mathbb{R}^2 : \mathbf{v}_\mathcal{T}\Big|_K \in R_1(K) \text{ for all } K \in \mathcal{T}, \text{ is continuous at mid-points, of interior edges, vanishes at mid-points of boundary edges} \Big\}$$

$$Y(\mathcal{T}) = S^{0,-1}(\mathcal{T}) \cap L^2_0(\Omega).$$

This pair of spaces is called *Crouzeix-Raviart element*. Its degrees of freedom are the velocity-vectors at the mid-points of interior edges and the pressures at the barycentres of elements. The corresponding discrete problem is:

Find $\mathbf{u}_\mathcal{T} \in X(\mathcal{T})$ and $p_\mathcal{T} \in Y(\mathcal{T})$ such that

$$\sum_{K\in\mathcal{T}} \int_K \nabla \mathbf{u}_\mathcal{T} : \nabla \mathbf{v}_\mathcal{T} \mathrm{d}x - \sum_{K\in\mathcal{T}} \int_K p_\mathcal{T} \operatorname{div} \mathbf{v}_\mathcal{T} \mathrm{d}x = \int_\Omega \mathbf{f} \cdot \mathbf{v}_\mathcal{T} \mathrm{d}x$$

$$\sum_{K\in\mathcal{T}} \int_K q_\mathcal{T} \operatorname{div} \mathbf{u}_\mathcal{T} \mathrm{d}x = 0$$

for all $\mathbf{v}_\mathcal{T} \in X(\mathcal{T})$ and all $q_\mathcal{T} \in Y(\mathcal{T})$.

The following results can be proven:

The Crouzeix-Raviart discretization admits a unique solution $\mathbf{u}_\mathcal{T}$, $p_\mathcal{T}$.

The continuity equation $\operatorname{div} \mathbf{u}_{\mathcal{T}} = 0$ is satisfied elementwise.
If Ω is *convex*, the following error estimates hold

$$\left\{ \sum_{K \in \mathcal{T}} |\mathbf{u} - \mathbf{u}_{\mathcal{T}}|^2_{H^1(K)} \right\}^{1/2} + \|p - p_{\mathcal{T}}\|_0 = \mathrm{O}(h),$$

$$\|\mathbf{u} - \mathbf{u}_{\mathcal{T}}\|_0 = \mathrm{O}(h^2).$$

Yet, the Crouzeix-Raviart discretization has several drawbacks too:

- Its accuracy deteriorates drastically in the presence of re-entrant corners.
- It has no higher order equivalent.

II.4.3. Construction of a local solenoidal bases. One of the main attractions of the Crouzeix-Raviart element is the fact that it allows the construction of a local solenoidal bases for the velocity space and that in this way the computation of the velocity and pressure can be decoupled.

For the construction of the solenoidal bases we assume that Ω is simply connected, ie. Γ has only one component, and denote by

- NT the number of triangles,
- NE_0 the number of interior edges,
- NV_0 the number of interior vertices,
- $V(\mathcal{T}) = \{\mathbf{u}_{\mathcal{T}} \in X(\mathcal{T}) : \operatorname{div} \mathbf{u}_{\mathcal{T}} = 0\}$ the space of solenoidal velocities.

The quantities NT, NE_0, and NV_0 are connected via *Euler's formula*

$$NT - NE_0 + NV_0 = 1.$$

Since $X(\mathcal{T})$ and $Y(\mathcal{T})$ satisfy the inf-sup condition, an elementary calculation using this formula yields

$$\begin{aligned} \dim V(\mathcal{T}) &= \dim X(\mathcal{T}) - \dim Y(\mathcal{T}) \\ &= 2NE_0 - (NT - 1) \\ &= NE_0 + NV_0. \end{aligned}$$

With every edge $E \in \mathcal{E}$ we associate a unit tangential vector $\mathbf{t}_E$ and a piecewise linear function φ_E which equals 1 at the midpoint of E and which vanishes at all other midpoints of edges. With this notation we set

$$\mathbf{w}_E = \varphi_E \mathbf{t}_E \quad \text{for all } E \in \mathcal{E}.$$

These functions obviously are linearly independent. For every triangle K and every edge E we have

$$\int_K \operatorname{div} \mathbf{w}_E \mathrm{d}x = \int_{\partial K} \mathbf{n}_K \cdot \mathbf{w}_E \mathrm{d}S = 0.$$

Since $\operatorname{div} \mathbf{w}_E$ is piecewise constant, this implies that the functions $\mathbf{w}_E$ are contained in $V(\mathcal{T})$.

With every vertex $x \in \mathcal{N}$ we associate the set $\mathcal{E}_x$ of all edges which have x as an endpoint. Starting with an arbitrary edge in $\mathcal{E}_x$ we enumerate the remaining edges in $\mathcal{E}_x$ consecutively in counter-clockwise orientation. For $E \in \mathcal{E}_x$ we denote by $\mathbf{t}_{E,x}$ and $\mathbf{n}_{E,x}$ two orthogonal unit vectors such that $\mathbf{t}_{E,x}$ is tangential to E, points away from x, and satisfies $\det(\mathbf{t}_{E,x}, \mathbf{n}_{E,x}) > 0$. With these notations we set (cf. Figure II.4.1)

$$\mathbf{w}_x = \sum_{E \in \mathcal{E}_x} \frac{1}{|E|} \varphi_E \mathbf{n}_{E,x}.$$

The functions $\mathbf{w}_x$ obviously are linearly independent. Since the $\mathbf{w}_E$ are tangential to the edges and the $\mathbf{w}_x$ are normal to the edges, both sets of functions are linearly independent.

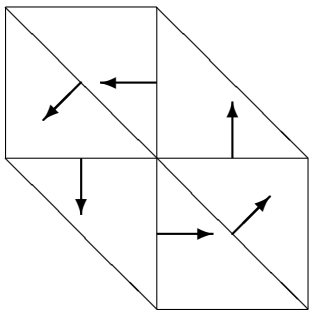

FIGURE II.4.1. The function $\mathbf{w}_x$

Next consider a triangle $K \in \mathcal{T}$ and a vertex x of K and denote by E_1, E_2 the two edges of K sharing the vertex x. Denote by m_{E_i} the midpoint of E_i, $i = 1, 2$. Using the integration by parts formulae of §I.2.3 (p. 18) and evaluating line integrals with the help of the midpoint formula, we conclude that

$$\begin{aligned}
\int_K \operatorname{div} \mathbf{w}_x \mathrm{d}x &= \int_{\partial K} \mathbf{w}_x \cdot \mathbf{n}_K \mathrm{d}S \\
&= \sum_{i=1}^{2} |E_i| \mathbf{w}_x(m_{E_i}) \cdot \mathbf{n}_K \\
&= \sum_{i=1}^{2} \mathbf{n}_{E_i,x} \cdot \mathbf{n}_K \\
&= 0.
\end{aligned}$$

This proves that $\mathbf{w}_x \in V(\mathcal{T})$.

Hence we have

$$\boxed{V(\mathcal{T}) = \operatorname{span}\{\mathbf{w}_x\,, \mathbf{w}_E\ : x \in \mathcal{N}\,, E \in \mathcal{E}\}}$$

and the discrete velocity field of the Crouzeix-Raviart discretization is given by the problem:

Find $\mathbf{u}_\mathcal{T} \in V(\mathcal{T})$ such that

$$\sum_{K\in\mathcal{T}} \int_K \nabla\mathbf{u}_\mathcal{T} : \nabla\mathbf{v}_\mathcal{T} \mathrm{d}x = \int_\Omega \mathbf{f}\cdot\mathbf{v}_\mathcal{T}\mathrm{d}x$$

for all $\mathbf{v}_\mathcal{T} \in V(\mathcal{T})$.

This is a linear system of equations with $NE_0 + NV_0$ equations and unknowns and with a symmetric, positive definite stiffness matrix. The stiffness matrix, however, has a condition of $\mathrm{O}(h^{-4})$ compared with a condition of $\mathrm{O}(h^{-2})$ of the larger, symmetric, but indefinite stiffness matrix of the original mixed problem.

The above problem only yields the velocity. The pressure, however, can be computed by a simple post-processing process. To describe it, denote by m_E the midpoint of any edge $E \in \mathcal{E}$. Using the integration by parts formulae of §I.2.3 (p. 18) and evaluating line integrals with the help of the midpoint formula, we conclude that

$$\begin{aligned}
&\int_\Omega \mathbf{f}\cdot\varphi_E\mathbf{n}_E\mathrm{d}x - \sum_{K\in\mathcal{T}}\int_K \nabla\mathbf{u}_\mathcal{T} : \nabla\varphi_E\mathbf{n}_E\mathrm{d}x\\
&\quad= -\sum_{K\in\mathcal{T}}\int_K p_\mathcal{T}\operatorname{div}\varphi_E\mathbf{n}_E\mathrm{d}x\\
&\quad= -\int_E \mathbb{J}_E(\varphi_E p_\mathcal{T})\mathrm{d}S\\
&\quad= -|E|\mathbb{J}_E(\varphi_E(m_E)p_\mathcal{T})\\
&\quad= -|E|\mathbb{J}_E(p_\mathcal{T}).
\end{aligned}$$

Hence, we can compute the pressure jumps edge by edge by evaluating the left-hand side of this equation. Starting with a triangle K which has an edge on the boundary Γ, we set $\widetilde{p}_\mathcal{T} = 0$ on this triangle and compute it on the remaining triangles by passing through adjacent triangles and adding the pressure jump corresponding to the common edge. Finally, we compute

$$P = \sum_{K\in\mathcal{T}_h} \frac{|K|}{|\Omega|}\widetilde{p}_\mathcal{T}\Big|_K$$

and subtract P from $\widetilde{p}_{\mathcal{T}}$. This yields the pressure $p_{\mathcal{T}}$ of the Crouzeix-Raviart discretization which has mean-value zero. This process is realized by Algorithm II.4.1.

Algorithm II.4.1 Pressure Computation

Require: $\mathbf{u}_{\mathcal{T}}$ solenoidal Crouzeix-Raviart solution
Provide: $p = p_{\mathcal{T}}$ corresponding pressure

1: $p \leftarrow 0$
2: choose an element $K \in \mathcal{T}$ having an edge on the boundary of Ω
3: $M \leftarrow K$, $\mathcal{U} \leftarrow \mathcal{E}_\Omega$
4: **while** $\mathcal{U} \neq \emptyset$ **do**
5: choose an edge $E \in \mathcal{U}$ with $\omega_E \cap M \neq \emptyset$
6: $p_{\omega_E \setminus M} \leftarrow p_{\omega_E \setminus M} + \int_{\omega_E} \nabla_{\mathcal{T}} \mathbf{u}_{\mathcal{T}} : \nabla_{\mathcal{T}}(\varphi_E \mathbf{n}_E) - \int_{\omega_E} \mathbf{f} \cdot (\varphi_E \mathbf{n}_E)$.
7: $M \leftarrow M \cup \omega_E$, $\mathcal{U} \leftarrow \mathcal{U} \setminus \{E\}$
8: **end while**
9: $p \leftarrow p - \sum_{K \in \mathcal{T}} \frac{|K|}{|\Omega|} p_K$.

II.5. Stream-function formulation

II.5.1. Motivation. The methods of the previous sections all discretize the variational formulation of the Stokes equations of §II.1.2 (p. 29) and yield simultaneously approximations for the velocity *and* pressure. The discrete velocity fields in general are not solenoidal, i.e. the incompressibility constraint is satisfied only approximately. In §II.4 we obtained an exactly solenoidal approximation but had to pay for this by abandoning the conformity. In this section we will consider another variational formulation of the Stokes equations which leads to conforming solenoidal discretizations. As we will see this advantage has to be paid for by other drawbacks.

Throughout this section we assume that Ω is a *two dimensional, simply connected polygonal domain.*

II.5.2. The curl operators. The subsequent analysis is based on two curl-operators which correspond to the rot-operator in three dimensions and which are defined as follows:

$$\mathbf{curl}\,\varphi = \begin{pmatrix} -\frac{\partial \varphi}{\partial y} \\ \frac{\partial \varphi}{\partial x} \end{pmatrix},$$
$$\operatorname{curl} \mathbf{v} = \frac{\partial v_1}{\partial y} - \frac{\partial v_2}{\partial x}.$$

They fulfill the following chain rule

$$\begin{aligned} &\operatorname{curl}(\mathbf{curl}\,\varphi) = -\Delta\varphi && \text{for all } \varphi \in H^2(\Omega) \\ &\mathbf{curl}(\operatorname{curl}\mathbf{v}) = -\Delta\mathbf{v} + \nabla(\operatorname{div}\mathbf{v}) && \text{for all } \mathbf{v} \in H^2(\Omega)^2 \end{aligned}$$

and the following integration by parts formula

$$\int_\Omega \mathbf{v}\cdot\mathbf{curl}\,\varphi \mathrm{d}x = \int_\Omega \operatorname{curl}\mathbf{v}\varphi \mathrm{d}x + \int_\Gamma \varphi\mathbf{v}\cdot\mathbf{t}\mathrm{d}S \quad \text{for all } \mathbf{v}\in H^1(\Omega)^2,\ \varphi\in H^1(\Omega).$$

Here $\mathbf{t}$ denotes a unit tangent vector to the boundary Γ.

The following deep mathematical result is fundamental:

A vector-field $\mathbf{v}:\Omega\to\mathbb{R}^2$ is solenoidal, i.e. $\operatorname{div}\mathbf{v}=0$, if and only if there is a unique *stream-function* $\varphi:\Omega\to\mathbb{R}$ such that $\mathbf{v}=\mathbf{curl}\,\varphi$ in Ω and $\varphi=0$ on Γ.

II.5.3. Stream-function formulation of the Stokes equations. Let $\mathbf{u}$, p be the solution of the Stokes equations with exterior force $\mathbf{f}$ and homogeneous boundary conditions and denote by ψ the stream function corresponding to $\mathbf{u}$. Since

$$\mathbf{u}\cdot\mathbf{t} = 0 \quad \text{on } \Gamma,$$

we conclude that in addition

$$\frac{\partial\psi}{\partial\mathbf{n}} = \mathbf{t}\cdot\mathbf{curl}\,\psi = 0 \quad \text{on } \Gamma.$$

Inserting this representation of $\mathbf{u}$ in the momentum equation and applying the operator curl we obtain

$$\begin{aligned} \operatorname{curl}\mathbf{f} &= \operatorname{curl}\{-\Delta\mathbf{u}+\nabla p\} \\ &= -\Delta(\operatorname{curl}\mathbf{u}) + \underbrace{\operatorname{curl}(\nabla p)}_{=0} \\ &= -\Delta(\operatorname{curl}(\mathbf{curl}\,\psi)) \\ &= \Delta^2\psi. \end{aligned}$$

This proves that the stream function ψ solves the *biharmonic equation*

$$\begin{aligned} \Delta^2\psi &= \operatorname{curl}\mathbf{f} && \text{in } \Omega \\ \psi &= 0 && \text{on } \Gamma \\ \frac{\partial\psi}{\partial\mathbf{n}} &= 0 && \text{on } \Gamma. \end{aligned}$$

Conversely, one can prove: If ψ solves the above biharmonic equation, there is a unique pressure p with mean-value 0 such that $\mathbf{u} = \mathbf{curl}\,\psi$ and p solve the Stokes equations. In this sense, the Stokes equations and the biharmonic equation are equivalent.

REMARK II.5.1. Given a solution ψ of the biharmonic equation and the corresponding velocity $\mathbf{u} = \mathbf{curl}\,\psi$ the pressure is determined by the equation $\mathbf{f} + \Delta\mathbf{u} = \nabla p$. Yet *there is no constructive way to solve this problem*. Hence, *the biharmonic equation is only capable to yield the velocity field of the Stokes equations*.

II.5.4. Variational formulation of the biharmonic equation. With the help of the Sobolev space

$$H_0^2(\Omega) = \left\{ \varphi \in H^2(\Omega) : \varphi = \frac{\partial \varphi}{\partial \mathbf{n}} = 0 \text{ on } \Gamma \right\}$$

the variational formulation of the biharmonic equation of the previous section is given by

Find $\psi \in H_0^2(\Omega)$ such that

$$\int_\Omega \Delta\psi \Delta\varphi \mathrm{d}x = \int_\Omega \operatorname{curl} \mathbf{f} \varphi \mathrm{d}x$$

for all $\varphi \in H_0^2(\Omega)$.

REMARK II.5.2. The above variational problem is the Euler-Lagrange equation corresponding to the minimization problem

$$J(\psi) = \frac{1}{2}\int_\Omega (\Delta\psi)^2 \mathrm{d}x - \int_\Omega \operatorname{curl} \mathbf{f}\psi \mathrm{d}x \to \min \quad \text{in } H_0^2(\Omega).$$

Therefore it is tempting to discretize the biharmonic equation by replacing in its variational formulation the space $H_0^2(\Omega)$ by a finite element space $X(\mathcal{T}) \subset H_0^2(\Omega)$. Yet, *this would require C^1-elements*! The *lowest polynomial degree* for which C^1-elements exist, *is five*!

II.5.5. A non-conforming discretization of the biharmonic equation. One possible remedy to the difficulties described in Remark II.5.2 is to drop the C^1-continuity of the finite element functions. This leads to non-conforming discretizations of the biharmonic equation. The most popular one is based on the so-called *Morley element*. It is a *triangular element*, i.e. the partition $\mathcal{T}$ exclusively consists of *triangles*. The corresponding finite element space is given by

$$M(\mathcal{T}) = \Big\{ \varphi \in S^{2,-1}(\mathcal{T}) : \varphi \text{ is continuous at the vertices,}$$

$$\left.\begin{array}{r}\mathbf{n}_E \cdot \nabla\varphi \text{ is continuous at} \\ \text{the midpoints of edges}\end{array}\right\}.$$

The degrees of freedom are

- the values of φ at the vertices and
- the values of $\mathbf{n}_E \cdot \nabla\varphi$ at the midpoints of edges.

The discrete problem is given by:

Find $\psi_{\mathcal{T}} \in M(\mathcal{T})$ such that

$$\sum_{K\in\mathcal{T}} \int_K \Delta\psi_{\mathcal{T}} \Delta\varphi_{\mathcal{T}} \mathrm{d}x = \sum_{K\in\mathcal{T}} \int_K \operatorname{curl} \mathbf{f}\, \varphi_{\mathcal{T}} \mathrm{d}x$$

for all $\varphi_{\mathcal{T}} \in M(\mathcal{T})$.

One can prove that this discrete problem admits a unique solution. The stiffness matrix is symmetric, positive definite and has a condition of $\mathrm{O}(h^{-4})$.

There is a close relation between this problem and the Crouzeix-Raviart discretization of §II.4 (p. 45):

If $\psi_{\mathcal{T}} \in M(\mathcal{T})$ is the solution of the Morley element discretization of the biharmonic equation, then $\mathbf{u}_{\mathcal{T}} = \mathbf{curl}\, \psi_{\mathcal{T}}$ is the solution of the Crouzeix-Raviart discretization of the Stokes equations.

II.5.6. Mixed finite element discretizations of the biharmonic equation. Another remedy to the difficulties described in Remark II.5.2 (p. 51) is to use a saddle-point formulation of the biharmonic equation and a corresponding mixed finite element discretization. This saddle-point formulation is obtained by introducing the *vorticity* $\omega = \operatorname{curl} \mathbf{u}$. It is given by:

- Find $\psi \in H^1_0(\Omega)$, $\omega \in L^2(\Omega)$ and $\lambda \in H^1(\Omega)$ such that

$$\int_\Omega \mathbf{curl}\, \lambda \cdot \mathbf{curl}\, \varphi \mathrm{d}x = \int_\Omega \mathbf{f} \cdot \mathbf{curl}\, \varphi \mathrm{d}x$$

$$\int_\Omega \omega\theta \mathrm{d}x - \int_\Omega \lambda\theta \mathrm{d}x = 0$$

$$\int_\Omega \mathbf{curl}\, \psi \cdot \mathbf{curl}\, \mu \mathrm{d}x - \int_\Omega \omega\mu \mathrm{d}x = 0$$

for all $\varphi \in H^1_0(\Omega)$, $\theta \in L^2(\Omega)$ and $\mu \in H^1(\Omega)$

- and find $p \in H^1(\Omega) \cap L^2_0(\Omega)$ such that

$$\int_\Omega \nabla p \cdot \nabla q \mathrm{d}x = \int_\Omega (\mathbf{f} - \mathbf{curl}\, \lambda) \cdot \nabla q \mathrm{d}x$$

for all $q \in H^1(\Omega) \cap L^2_0(\Omega)$.

REMARK II.5.3. Compared with the saddle-point formulation of the Stokes equations of §II.1.2 (p. 29), the above problem imposes weaker regularity conditions on the velocity $\mathbf{u} = \mathbf{curl}\, \psi$ and stronger regularity conditions on the pressure p. The two quantities ω and λ are both vorticities. The second equation means that they are equal in a weak sense. Yet, this equality is not a point-wise one since both quantities have different regularity properties. The computation of the pressure is decoupled from the computation of the other quantities.

For the mixed finite element discretization we depart from a *triangulation* $\mathcal{T}$ of Ω and choose an integer $\ell \geq 1$ and set $k = \max\{1, \ell - 1\}$. Then the discrete problem is given by:

Find $\psi_\mathcal{T} \in S^{\ell,0}_0(\mathcal{T})$, $\omega_\mathcal{T} \in S^{\ell,-1}(\mathcal{T})$ and $\lambda_\mathcal{T} \in S^{\ell,0}(\mathcal{T})$ such that

$$\int_\Omega \mathbf{curl}\, \lambda_\mathcal{T} \cdot \mathbf{curl}\, \varphi_\mathcal{T} \mathrm{d}x = \int_\Omega \mathbf{f} \cdot \mathbf{curl}\, \varphi_\mathcal{T} \mathrm{d}x$$

$$\int_\Omega \omega_\mathcal{T} \theta_\mathcal{T} \mathrm{d}x - \int_\Omega \lambda_\mathcal{T} \theta_\mathcal{T} \mathrm{d}x = 0$$

$$\int_\Omega \mathbf{curl}\, \psi_\mathcal{T} \cdot \mathbf{curl}\, \mu_\mathcal{T} \mathrm{d}x - \int_\Omega \omega_\mathcal{T} \mu_\mathcal{T} \mathrm{d}x = 0$$

for all $\varphi_\mathcal{T} \in S^{\ell,0}_0(\mathcal{T})$, $\theta_\mathcal{T} \in S^{\ell,-1}(\mathcal{T})$ and $\mu_\mathcal{T} \in S^{\ell,0}(\mathcal{T})$ and find $p_\mathcal{T} \in S^{k,0}(\mathcal{T}) \cap L^2_0(\Omega)$ such that

$$\int_\Omega \nabla p_\mathcal{T} \cdot \nabla q_\mathcal{T} \mathrm{d}x = \int_\Omega (\mathbf{f} - \mathbf{curl}\, \lambda_\mathcal{T}) \cdot \nabla q_\mathcal{T} \mathrm{d}x$$

for all $q_\mathcal{T} \in S^{k,0}(\mathcal{T}) \cap L^2_0(\Omega)$

One can prove that this discrete problem is well-posed and that its solution satisfies the following a priori error estimate

$$\begin{aligned} &\|\psi - \psi_\mathcal{T}\|_1 + \|\omega - \omega_\mathcal{T}\|_0 + \|p - p_\mathcal{T}\|_0 \\ &\leq c \left\{ h^{m-1} \left[\|\psi\|_{m+1} + \|\Delta\psi\|_m + \|p\|_{\max\{1,m-1\}} \right] + h^m \|\Delta\psi\|_m \right\}, \end{aligned}$$

where $1 \leq m \leq \ell$ depends on the regularity of the solution of the biharmonic equation. In the lowest order case $\ell = 1$ this estimate does not imply convergence. Yet, using a more sophisticated analysis, one

can prove in this case an $O(h^{\frac{1}{2}})$-error estimate provided $\psi \in H^3(\Omega)$, i.e. $\mathbf{u} \in H^2(\Omega)^2$.

II.6. Solution of the discrete problems

II.6.1. General structure of the discrete problems. For the discretization of the Stokes equations we choose a partition $\mathcal{T}$ of the domain Ω and finite element spaces $X(\mathcal{T})$ and $Y(\mathcal{T})$ for the approximation of the velocity and the pressure, respectively. With these choices we either consider a mixed method as in §II.2 (p. 32) or a Petrov-Galerkin one as in §II.3 (p. 40). Denote by

- $n_{\mathbf{u}}$ the dimension of $X(\mathcal{T})$ and by
- n_p the dimension of $Y(\mathcal{T})$ respectively.

Then the discrete problem has the form

$$\begin{pmatrix} A & B \\ B^T & -\delta C \end{pmatrix} \begin{pmatrix} \mathbf{u} \\ p \end{pmatrix} = \begin{pmatrix} \mathbf{f} \\ \delta g \end{pmatrix} \tag{II.6.1}$$

with

- $\delta = 0$ for the mixed methods of §II.2,
- $\delta > 0$ and $\delta \approx 1$ for the Petrov-Galerkin methods of §II.3,
- a square, symmetrix, positive definite $n_{\mathbf{u}} \times n_{\mathbf{u}}$ matrix A with condition of $O(h^{-2})$,
- a rectangular $n_{\mathbf{u}} \times n_p$ matrix B,
- a square, symmetric, positive definite $n_p \times n_p$ matrix C with condition of $O(1)$,
- a vector $\mathbf{f}$ of dimension $n_{\mathbf{u}}$ discretizing the exterior force, and
- a vector g of dimension n_p which equals 0 for the mixed methods of §II.2 and which results from the stabilization terms on the right-hand sides of the methods of §II.3.

Remark II.6.1. For simplicity, we drop in this section the index $\mathcal{T}$ indicating the discretization. Moreover, we identify finite element functions with their coefficient vectors with respect to a nodal bases. Thus, $\mathbf{u}$ and p are now vectors of dimension $n_{\mathbf{u}}$ and n_p, respectively.

The stiffness matrix

$$\begin{pmatrix} A & B \\ B^T & -\delta C \end{pmatrix}$$

of the discrete problem (II.6.1) is symmetric, but *indefinite*, i.e. it has positive *and* negative real eigenvalues. Correspondingly, well-established methods such as the conjugate gradient algorithm cannot be applied. This is the main difficulty in solving the linear systems arising from the discretization of the Stokes equations.

Remark II.6.2. The indefiniteness of the stiffness matrix results from the saddle-point structure of the Stokes equations. This cannot be remedied by any stabilization procedure.

II.6.2. The Uzawa algorithm. The Uzawa algorithm II.6.1 probably is the simplest algorithm for solving the discrete problem (II.6.1).

Algorithm II.6.1 Uzawa algorithm

Require: initial guess p, tolerance $\varepsilon > 0$, relaxation parameter $\omega > 0$, maximal number of iterations N.
Provide: approximate solution $\mathbf{u}$, p of (II.6.1).

1: $E \leftarrow \infty$, $n \leftarrow 0$
2: **while** $E > \varepsilon$ and $n \leq N$ **do**
3: Apply a few Gauß-Seidel iterations to $A\mathbf{u} = \mathbf{f} - Bp$; result $\mathbf{u}$.
4: $p \leftarrow p + \omega\{B^T\mathbf{u} - \delta g - \delta Cp\}$
5: $E \leftarrow \|A\mathbf{u} + Bp - \mathbf{f}\| + \|B^T\mathbf{u} - \delta Cp - \delta g\|$, $n \leftarrow n + 1$
6: **end while**

Remark II.6.3. *(1)* The relaxation parameter ω usually is chosen in the interval $(1, 2)$; a typical choice is $\omega = 1.5$.
(2) $\|\cdot\|$ denotes any vector norm. A popular choice is the scaled Euclidean norm, i.e. $\|\mathbf{v}\| = \sqrt{\frac{1}{n_\mathbf{u}}\mathbf{v}\cdot\mathbf{v}}$ for the velocity part and $\|q\| = \sqrt{\frac{1}{n_p}q\cdot q}$ for the pressure part.
(3) The problem $A\mathbf{u} = \mathbf{f} - Bp_i$ is a discrete version of two, if $\Omega \subset \mathbb{R}^2$, or three, if $\Omega \subset \mathbb{R}^3$, Poisson equations for the components of the velocity field.
(4) The Uzawa algorithm iterates on the pressure. Therefore it is sometimes called a *pressure correction scheme*.

The Uzawa algorithm is very simple, but extremely slow. Therefore it cannot be recommended for practical use. We have given it nevertheless, since it is the basis for the more efficient algorithm of §II.6.4 (p. 56).

II.6.3. The conjugate gradient algorithm revisited. The algorithm of the next section is based on the conjugate gradient algorithm (CG algorithm) II.6.2.

Algorithm II.6.2 Conjugate gradient algorithm

Require: matrix L, right-hand side b, initial guess x, tolerance ε, maximal number of iterations N.
Provide: approximate solution x with $\|Lx - b\| \leq \varepsilon$.

1: $r \leftarrow b - Lx$, $d \leftarrow r$, $\gamma \leftarrow r \cdot r$, $n \leftarrow 0$
2: **while** $\gamma > \varepsilon^2$ und $n \leq N$ **do**
3: $s \leftarrow Ld$, $\alpha \leftarrow \frac{\gamma}{d\cdot s}$, $x \leftarrow x + \alpha d$, $r \leftarrow r - \alpha s$
4: $\beta \leftarrow \frac{r\cdot r}{\gamma}$, $\gamma \leftarrow r \cdot r$, $d \leftarrow r + \beta d$, $n \leftarrow n + 1$
5: **end while**

The convergence rate of the CG-algorithm is given by $\frac{(\sqrt{\kappa}-1)}{(\sqrt{\kappa}+1)}$ where κ is the condition of L and equals the ratio of the largest to the smallest eigenvalue of L.

II.6.4. An improved Uzawa algorithm. Since the matrix A is positive definite, we may solve the first equation in the system (II.6.1) for the unknown $\mathbf{u}$

$$\mathbf{u} = A^{-1}(\mathbf{f} - Bp)$$

and insert the result in the second equation

$$B^T A^{-1}(\mathbf{f} - Bp) - \delta Cp = \delta g.$$

This gives a problem which only incorporates the pressure

$$[B^T A^{-1} B + \delta C]p = B^T A^{-1}\mathbf{f} - \delta g. \tag{II.6.2}$$

One can prove that the matrix $B^T A^{-1} B + \delta C$ is symmetric, positive definite and has a condition of O(1), i.e. the condition does not increase when refining the mesh. Therefore one may apply the CG-algorithm to problem (II.6.2). The convergence rate then is independent of the mesh-size and does not deteriorate when refining the mesh. This approach, however, requires the evaluation of A^{-1}, i.e. problems of the form $A\mathbf{v} = \mathbf{g}$ must be solved in every iteration. These are two, if $\Omega \subset \mathbb{R}^2$, or three, if $\Omega \subset \mathbb{R}^3$, discrete Poisson equations for the components of the velocity $\mathbf{v}$. The crucial idea now is to solve these auxiliary problems only approximately with the help of a standard multigrid algorithm for the Poisson equation.

This idea results in Algorithm II.6.3.

Algorithm II.6.3 Improved Uzawa algorithm

Require: initial guess p, tolerance $\varepsilon > 0$, maximal number of iterations N.

Provide: approximate solution $\mathbf{u}$, p of (II.6.1).

1: Apply a multigrid algorithm with starting value zero and tolerance ε to $A\mathbf{v} = \mathbf{f} - Bp$; result $\mathbf{u}$.
2: $r \leftarrow B^T\mathbf{u} - \delta g - \delta Cp$, $d \leftarrow r$, $\gamma \leftarrow r \cdot r$
3: $\mathbf{u} \leftarrow 0$, $q \leftarrow p$, $p \leftarrow 0$, $n \leftarrow 0$
4: **while** $\gamma > \varepsilon^2$ and $n \leq N$ **do**
5: Apply a multigrid algorithm with starting value $\mathbf{u}$ and tolerance ε to $A\mathbf{v} = Bd$; result $\mathbf{u}$.
6: $s \leftarrow B^T\mathbf{u} + \delta Cd$, $\alpha \leftarrow \frac{\gamma}{d \cdot s}$, $p \leftarrow p + \alpha d$, $r \leftarrow r - \alpha s$
7: $\beta \leftarrow \frac{r \cdot r}{\gamma}$, $\gamma \leftarrow r \cdot r$, $d \leftarrow r + \beta d$, $n \leftarrow n + 1$
8: **end while**
9: $p \leftarrow q + p$
10: Apply a multigrid algorithm with starting value zero and tolerance ε to $A\mathbf{v} = \mathbf{f} - Bp$; result $\mathbf{u}$.

REMARK II.6.4. The improved Uzawa algorithm is a nested iteration: The outer iteration is a CG-algorithm for problem (II.6.2), the inner iteration is a standard multigrid algorithm for discrete Poisson equations. The convergence rate of the improved Uzawa algorithm does not deteriorate when refining the mesh. It usually lies in the range of 0.5 to 0.8. For the inner loop usually 2 to 4 multigrid iterations are sufficient.

II.6.5. The multigrid algorithm. The multigrid algorithm is based on a sequence of meshes $\mathcal{T}_0, \ldots, \mathcal{T}_R$, which are obtained by successive local or global refinement, and associated discrete problems $L_k x_k = b_k$, $k = 0, \ldots, R$, corresponding to a partial differential equation. The finest mesh $\mathcal{T}_R$ corresponds to the problem that we actually want to solve. In our applications, the differential equation is either the Stokes problem or the Poisson equation. In the first case L_k is the stiffness matrix of problem (II.6.1) corresponding to the partition $\mathcal{T}_k$. The vector x_k then incorporates the velocity and the pressure approximation. In the second case L_k is the upper left block A in problem (II.6.1) and x_k only incorporates the discrete velocity.

The multigrid algorithm has three ingredients:

- a *smoothing operator* M_k, which should be easy to evaluate and which at the same time should give a reasonable approximation to L_k^{-1},
- a *restriction operator* $R_{k,k-1}$, which maps functions on a fine mesh $\mathcal{T}_k$ to the next coarser mesh $\mathcal{T}_{k-1}$, and
- a *prolongation operator* $I_{k-1,k}$, which maps functions from a coarse mesh $\mathcal{T}_{k-1}$ to the next finer mesh $\mathcal{T}_k$.

For a concrete multigrid algorithm these ingredients must be specified. This will be done in the next sections. Here, we discuss the general form II.6.4 of the algorithm and its properties.

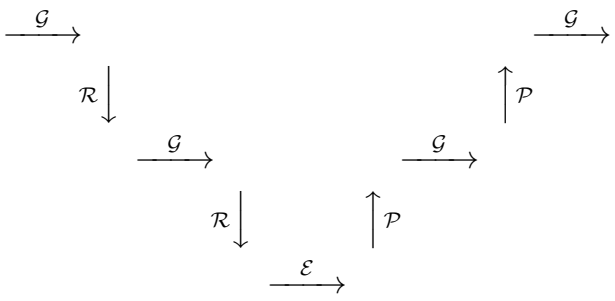

FIGURE II.6.1. Schematic presentation of a multigrid algorithm with V-cycle and three grids. The labels have the following meaning: $\mathcal{S}$ smoothing, $\mathcal{R}$ restriction, $\mathcal{P}$ prolongation, $\mathcal{E}$ exact solution.

Algorithm II.6.4 $\texttt{MG}(k, \mu, \nu_1, \nu_2, L_k, b, x)$ one multigrid iteration on mesh $\mathcal{T}_k$

Require: level number k, parameters μ, ν_1, ν_2, stiffness matrix L_k, right-hand side b, approximation M_k for L_k^{-1}, initial guess x.
Provide: improved approximate solution x.

1: **if** $k = 0$ **then**
2: $\quad x \leftarrow L_0^{-1} b$, `stop`
3: **end if**
4: **for** $i = 1, \ldots, \nu_1$ **do** ▷ Pre-smoothing
5: $\quad x \leftarrow x + M_k(b - L_k x)$
6: **end for**
7: $f \leftarrow R_{k,k-1}(b - L_k x)$, $y \leftarrow 0$ ▷ Coarse grid correction
8: Perform μ iterations of $\texttt{MG}(k-1, \mu, \nu_1, \nu_2, L_{k-1}, f, y)$; result y.
9: $x \leftarrow x + I_{k-1,k} y$
10: **for** $i = 1, \ldots, \nu_2$ **do** ▷ Post-smoothing
11: $\quad x \leftarrow x + M_k(b - L_k x)$
12: **end for**

REMARK II.6.5. *(1)* The parameter μ determines the complexity of the algorithm. Popular choices are $\mu = 1$ called *V-cycle* and $\mu = 2$ called *W-cycle*. Figure II.6.1 gives a schematic presentation of the multigrid algorithm for the case $\mu = 1$ and $R = 2$ (three meshes). Here, $\mathcal{S}$ denotes smoothing, $\mathcal{R}$ restriction, $\mathcal{P}$ prolongation, and $\mathcal{E}$ exact solution.
(2) The number of smoothing steps per multigrid iteration, i.e. the parameters ν_1 and ν_2, should not be chosen too large. A good choice for positive definite problems such as the Poisson equation is $\nu_1 = \nu_2 = 1$. For indefinite problems such as the Stokes equations a good choice is $\nu_1 = \nu_2 = 2$.
(3) If $\mu \leq 2$, one can prove that the computational work of one multigrid iteration is proportional to the number of unknowns of the actual discrete problem.
(4) Under suitable conditions on the smoothing algorithm, which is determined by the matrix M_k, one can prove that the convergence rate of the multigrid algorithm is independent of the mesh-size, i.e. it does not deteriorate when refining the mesh. These conditions will be discussed in the next section. In practice one observes convergence rates of 0.1 – 0.5 for positive definite problems such as the Poisson equation and of 0.3 – 0.7 for indefinite problems such as the Stokes equations.

II.6.6. Smoothing. The symmetric *Gauß-Seidel algorithm* is the most popular smoothing algorithm for positive definite problems such as the Poisson equation. This corresponds to the choice

$$M_k = (D_k - U_k^T) D_k^{-1} (D_k - U_k),$$

where D_k and U_k denote the diagonal and the strictly upper diagonal part of L_k respectively.

For indefinite problems such as the Stokes equations the most popular smoothing algorithm is the squared Jacobi iteration. This is the Jacobi iteration applied to the squared system $L_k^T L_k x_k = L_k^T b_k$ and corresponds to the choice

$$M_k = \omega^{-2} L_k^T$$

with a suitable damping parameter satisfying $\omega > 0$ and $\omega = \mathrm{O}(h_K^{-2})$.

Another class of popular smoothing algorithms for the Stokes equations is given by the so-called *Vanka methods*. The idea is to loop through patches of elements and to solve exactly the equations associated with the nodes inside the actual patch while retaining the current values of the variables associated with the nodes outside the actual patch.

There are many possible choices for the patches.

One extreme case obviously consists in choosing exactly one node at a time. This yields the classical Gauß-Seidel method and is not applicable to indefinite problems, since it in general diverges for those problems.

Another extreme case obviously consists in choosing all elements. This of course is not practicable since it would result in an exact solution of the complete discrete problem.

In practice, one chooses patches that consist of

- a single element, or
- the two elements that share a given edge or face, or
- the elements that share a given vertex.

II.6.7. Prolongation. Since the partition $\mathcal{T}_k$ of level k always is a refinement of the partition $\mathcal{T}_{k-1}$ of level $k-1$ (cf. §§II.7.6 (p. 70), II.7.7 (p. 70)), the corresponding finite element spaces are nested, i.e. finite element functions corresponding to level $k-1$ are contained in the finite element space corresponding to level k. Therefore, the values of a coarse-grid function corresponding to level $k-1$ at the nodal points corresponding to level k are obtained by evaluating the nodal bases functions corresponding to $\mathcal{T}_{k-1}$ at the requested points. This defines the interpolation operator $I_{k-1,k}$.

Figures II.6.2 and II.6.3 show various partitions of a triangle and of a square, respectively (cf. §II.7.6 (p. 70), II.7.7 (p. 70)). The numbers outside the element indicate the enumeration of the element vertices and edges. Thus, e.g. edge 2 of the triangle has the vertices 0 and 1 as its endpoints. The numbers +0, +1 etc. inside the elements indicate the enumeration of the child elements. The remaining numbers inside the elements give the enumeration of the vertices of the child elements.

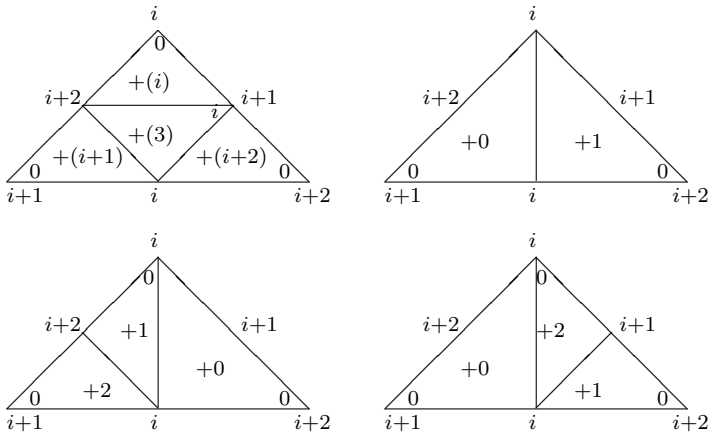

FIGURE II.6.2. Partitions of a triangle; expressions of the form $i + 1$ have to be taken modulo 3. The numbers outside the element indicate the enumeration of the element vertices and edges. The numbers +0, +1 etc. inside the elements indicate the enumeration of the child elements.

EXAMPLE II.6.6. Consider a piecewise constant approximation, i.e. $S^{0,-1}(\mathcal{T})$. The nodal points are the barycentres of the elements. Every element in $\mathcal{T}_{k-1}$ is subdivided into several smaller elements in $\mathcal{T}_k$. The nodal value of a coarse-grid function at the barycentre of a child element in $\mathcal{T}_k$ then is its nodal value at the barycentre of the parent element in $\mathcal{T}_k$.

EXAMPLE II.6.7. Consider a piecewise linear approximation, i.e. $S^{1,0}(\mathcal{T})$. The nodal points are the vertices of the elements. The refinement introduces new vertices at the midpoints of some edges of the parent element and possibly, when using quadrilaterals, at the barycentre of the parent element. The nodal value at the midpoint of an edge is the average of the nodal values at the endpoints of the edge. Thus, e.g. the value at vertex 1 of child +0 is the average of the values at vertices 0 and 1 of the parent element. Similarly, the nodal value at the barycentre of the parent element is the average of the nodal values at the four element vertices.

II.6.8. Restriction. The restriction is computed by expressing the nodal bases functions corresponding to the coarse partition $\mathcal{T}_{k-1}$ in terms of the nodal bases functions corresponding to the fine partition $\mathcal{T}_k$ and inserting this expression in the variational formulation. This results in a lumping of the right-hand side vector which, in a certain sense, is the transpose of the interpolation.

EXAMPLE II.6.8. Consider a piecewise constant approximation, i.e. $S^{0,-1}(\mathcal{T})$. The nodal shape function of a parent element is the sum of

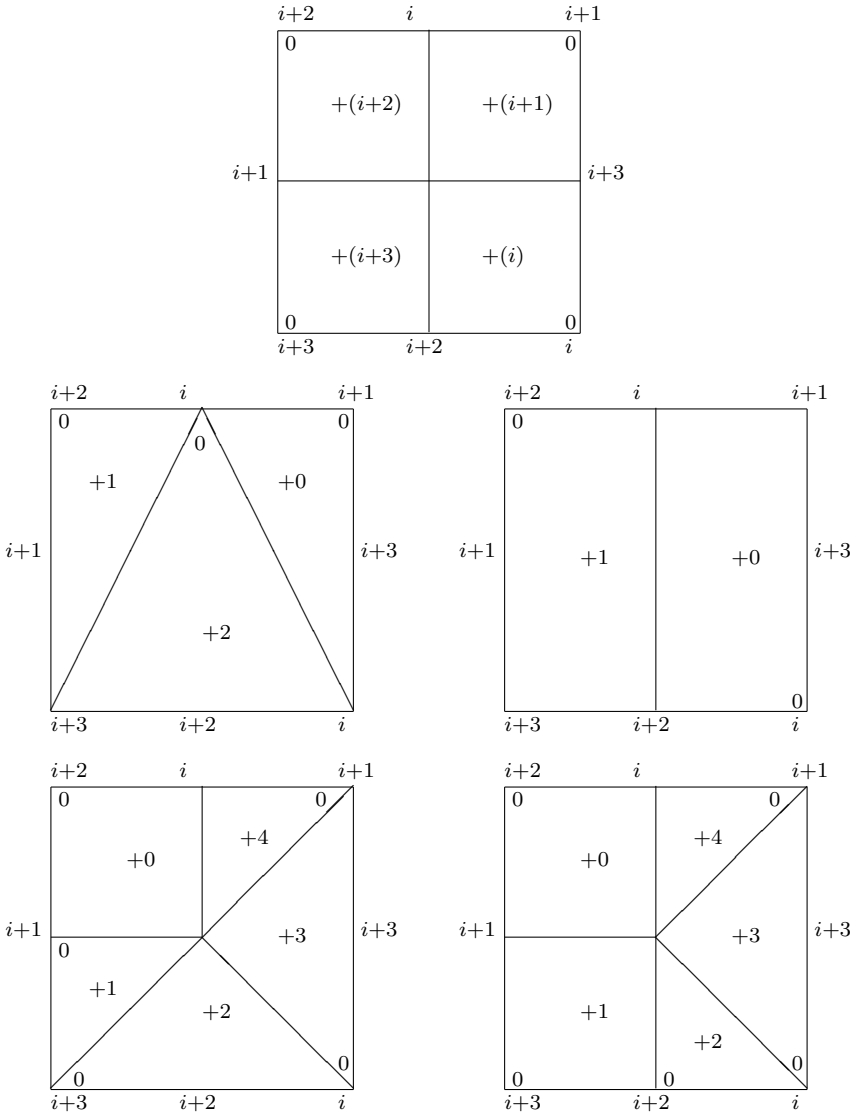

FIGURE II.6.3. Partitions of a square; expressions of the form $i+1$ have to be taken modulo 4. The numbers outside the element indicate the enumeration of the element vertices and edges. The numbers +0, +1 etc. inside the elements indicate the enumeration of the child elements.

the nodal shape functions of the child elements. Correspondingly, the components of the right-hand side vector corresponding to the child elements are all added and associated with the parent element.

EXAMPLE II.6.9. Consider a piecewise linear approximation, i.e. $S^{1,0}(\mathcal{T})$. The nodal shape function corresponding to a vertex of a parent *triangle* takes the value 1 at this vertex, the value $\frac{1}{2}$ at the midpoints of the two edges sharing the given vertex and the value 0

on the remaining edges. If we label the current vertex by a and the midpoints of the two edges emanating form a by m_1 and m_2, this results in the following formula for the restriction on a *triangle*

$$R_{k,k-1}\psi(a) = \psi(a) + \frac{1}{2}\{\psi(m_1) + \psi(m_2)\}.$$

When considering a *quadrilateral*, we must take into account that the nodal shape functions take the value $\frac{1}{4}$ at the barycentre b of the parent quadrilateral. Therefore the restriction on a *quadrilateral* is given by the formula

$$R_{k,k-1}\psi(a) = \psi(a) + \frac{1}{2}\{\psi(m_1) + \psi(m_2)\} + \frac{1}{4}\psi(b).$$

REMARK II.6.10. An efficient implementation of the prolongation and restrictions loops through all elements and performs the prolongation or restriction element-wise. This process is similar to the usual element-wise assembly of the stiffness matrix and the load vector.

II.6.9. Variants of the CG-algorithm for indefinite problems. The CG-algorithm can only be applied to symmetric positive definite systems of equations. For non-symmetric or indefinite systems it in general breaks down. Yet, there are various variants of the CG-algorithm which can be applied to these problems.

A naive approach consists in applying the CG-algorithm to the squared system $L_k^T L_k x_k = L_k^T b_k$ which is symmetric and positive definite. This approach cannot be recommended since squaring the systems squares its condition number and thus at least doubles the required number of iterations.

A more efficient algorithm is the *stabilized bi-conjugate gradient algorithm* II.6.5, shortly *Bi-CG-stab*. The underlying idea roughly is to solve simultaneously the original problem $L_k x_k = b_k$ and its adjoint $L_k^T y_k = b_k^T$.

II.7. A posteriori error estimation and adaptive grid refinement

II.7.1. Motivation. Suppose we have computed the solution of a discretization of a partial differential equation such as the Stokes equations. What is the error of our computed solution?

A priori error estimates do not help in answering this question. They only describe the asymptotic behaviour of the error. They tell us how fast it will converge to zero when refining the underlying mesh. Yet for a given mesh and discretization they give no information on the actual size of the error.

Another closely related problem is the question about the spatial distribution of the error. Where is it large, where is it small? Obviously we want to concentrate our resources in areas of a large error.

Algorithm II.6.5 Stabilized bi-conjugate gradient algorithm Bi-CG-stab

Require: matrix L, right-hand side b, initial guess x, tolerance ε, maximal number of iterations N.
Provide: approximate solution x with $\|Lx - b\| \leq \varepsilon$.

$r \leftarrow b - Lx$, $n \leftarrow 0$, $\gamma \leftarrow r \cdot r$
$\overline{r} \leftarrow r$, $\widehat{r} \leftarrow r$, $v \leftarrow 0$, $p \leftarrow 0$, $\alpha \leftarrow 1$, $\rho \leftarrow 1$, $\omega \leftarrow 1$
while $\gamma > \varepsilon^2$ and $n \leq N$ **do**
    $\beta \leftarrow \frac{\overline{r} \cdot r \alpha}{\rho \omega}$, $\rho \leftarrow \overline{r} \cdot r$
    if $|\beta| < \varepsilon$ **then**
        `stop` ▷ Break-down
    end if
    $p \leftarrow r + \beta\{p - \omega v\}$, $v \leftarrow Lp$, $\alpha \leftarrow \frac{\rho}{\widehat{r} \cdot v}$
    if $|\alpha| < \varepsilon$ **then**
        `stop` ▷ Break-down
    end if
    $s \leftarrow r - \alpha v$, $t \leftarrow Ls$, $\omega \leftarrow \frac{t \cdot s}{t \cdot t}$
    $x \leftarrow x + \alpha p + \omega s$, $r \leftarrow s - \omega t$, $n \leftarrow n + 1$
end while

In summary, we want to compute an approximate solution of the partial differential equation with a given tolerance and a minimal amount of work. This task is achieved by adaptive grid refinement based on a posteriori error estimation.

Throughout this section we denote by $X(\mathcal{T})$ and $Y(\mathcal{T})$ finite element spaces for the velocity and pressure, respectively associated with a given partition $\mathcal{T}$ of the domain Ω. With these spaces we associate either a mixed method as in §II.2 (p. 32) or a Petrov-Galerkin method as in §II.3 (p. 40). The solution of the corresponding discrete problem is denoted by $\mathbf{u}_\mathcal{T}$, $p_\mathcal{T}$, whereas $\mathbf{u}$, p denotes the solution of the Stokes equations.

II.7.2. General structure of the adaptive algorithm. The adaptive algorithm II.7.1 has a general structure which is independent of the particular differential equation.

In order to make the adaptive algorithm operative we must obviously specify the following ingredients:

- an algorithm that computes the η_K's (*a posteriori error estimation*),
- a rule that selects the elements in $\widetilde{\mathcal{T}}_k$ (*marking strategy*),
- a rule that refines the elements in $\widetilde{\mathcal{T}}_k$ (*regular refinement*),
- an algorithm that constructs the partition $\mathcal{T}_{k+1}$ (*additional refinement*).

Algorithm II.7.1 General adaptive algorithm

Require: data of the pde, tolerance ε.
Provide: approximate solution to the pde with error less than ε.
Construct an initial admissible partition $\mathcal{T}_0$.
for $k = 0, 1, \ldots$ **do**
  Solve the discrete problem corresponding to $\mathcal{T}_k$.
  for $K \in \mathcal{T}_k$ **do**
    Compute an estimate η_K of the error on K.
  end for
  $\eta \leftarrow \left\{\sum_{K\in\mathcal{T}_K} \eta_K^2\right\}^{1/2}$
  if $\eta \leq \varepsilon$ **then**
    `stop` ▷ Desired accuracy attained
  end if
  Based on $(\eta_K)_K$ determine a set $\widetilde{\mathcal{T}}_k$ of elements to be refined.
  Based on $\widetilde{\mathcal{T}}_k$ determine an admissible refinement $\mathcal{T}_{k+1}$ of $\mathcal{T}_k$.
end for

REMARK II.7.1. The η_K's are usually called *(a posteriori) error estimators* or *(a posteriori) error indicators*.

II.7.3. A residual a posteriori error estimator. The simplest and most popular a posteriori error estimator is the *residual error estimator*. For the Stokes equations it is given by

$$\eta_K = \left\{ h_K^2 \|\mathbf{f} + \Delta \mathbf{u}_\mathcal{T} - \nabla p_\mathcal{T}\|_{L^2(K)}^2 + \|\operatorname{div} \mathbf{u}_\mathcal{T}\|_{L^2(K)}^2 + \frac{1}{2} \sum_{E\in\mathcal{E}_K} h_E \|\mathbb{J}_E(\mathbf{n}_E \cdot (\nabla \mathbf{u}_\mathcal{T} - p_\mathcal{T} \underline{\mathbf{I}}))\|_{L^2(E)}^2 \right\}^{1/2},$$

where $\mathcal{E}_K$ is the collection of all edges or faces of K. Note that:

- The first term is the residual of the discrete solution with respect to the momentum equation $-\Delta \mathbf{u} + \nabla p = \mathbf{f}$ in its strong form.
- The second term is the residual of the discrete solution with respect to the continuity equation $\operatorname{div} \mathbf{u} = 0$ in it strong form.
- The third term contains the boundary terms that appear when switching from the strong to the weak form of the momentum equation by using an integration by parts formula elementwise.
- The pressure jumps vanish when using a continuous pressure approximation.

One can prove that the residual error estimator yields the following upper and lower bounds on the error:

$$\|\mathbf{u}-\mathbf{u}_{\mathcal{T}}\|_1+\|p-p_{\mathcal{T}}\|_0 \le c^*\left\{\sum_{K\in\mathcal{T}}\eta_K^2\right\}^{1/2},$$
$$\eta_K \le c_*\Big\{\|\mathbf{u}-\mathbf{u}_{\mathcal{T}}\|_{H^1(\omega_K)}+\|p-p_{\mathcal{T}}\|_{L^2(\omega_K)} + h_K\|\mathbf{f}-\mathbf{f}_{\mathcal{T}}\|_{L^2(\omega_K)}\Big\}.$$

Here, $\mathbf{f}_{\mathcal{T}}$ is the L^2-projection of $\mathbf{f}$ onto $S^{0,-1}(\mathcal{T})$ and ω_K denotes the union of all elements that share an edge with K, if $\Omega\subset\mathbb{R}^2$, or a face, if $\Omega\subset\mathbb{R}^3$ (cf. Figure II.7.1). The $\mathbf{f}-\mathbf{f}_{\mathcal{T}}$-term often is of higher order. The constant c^* depends on the constant c_Ω in the stability result at the end of §II.2.1 (p. 32) and on the shape parameter $\max_{K\in\mathcal{T}}\frac{h_K}{\rho_K}$ of the partition $\mathcal{T}$. The constant c_* depends on the shape parameter of $\mathcal{T}$ and on the polynomial degree of the spaces $X(\mathcal{T})$ and $Y(\mathcal{T})$.

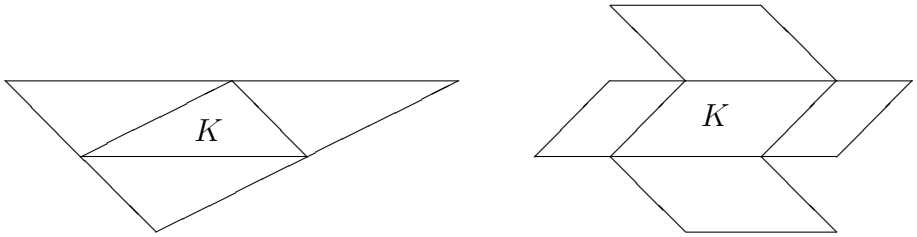

FIGURE II.7.1. Domain ω_K for a triangle and a parallelogram

REMARK II.7.2. The lower bound on the error is a local one whereas the upper bound is a global one. This is not by chance. The lower error bound involves the differential operator which is a local one: local variations in the velocity and pressure result in local variations of the force terms. The upper error bound on the other hand involves the inverse of the differential operator which is a global one: local variations of the exterior forces result in a global change of the velocity and pressure.

REMARK II.7.3. An error estimator, which yields an upper bound on the error, is called *reliable*. An estimator, which yields a lower bound on the error, is called *efficient*. Any decent error estimator must be reliable and efficient.

II.7.4. Error estimators based on the solution of auxiliary problems. Two other classes of popular error estimators are based on the solution of auxiliary discrete local problems with Neumann and

Dirichlet boundary conditions. For their description we denote by $k_{\mathbf{u}}$ and k_p the maximal polynomial degrees k and m such that

- $S_0^{k,0}(\mathcal{T})^n \subset X(\mathcal{T})$

and either

- $S^{m,-1}(\mathcal{T}) \cap L_0^2(\Omega) \subset Y(\mathcal{T})$ when using a discontinuous pressure approximation or
- $S^{m,0}(\mathcal{T}) \cap L_0^2(\Omega) \subset Y(\mathcal{T})$ when using a continuous pressure approximation.

Set

$$k_{\mathcal{T}} = \max\{k_{\mathbf{u}} + n, k_p - 1\},$$
$$k_{\mathcal{E}} = \max\{k_{\mathbf{u}} - 1, k_p\},$$

where n is the space dimension, i.e. $\Omega \subset \mathbb{R}^n$.

For the Neumann-type estimator we introduce for every element $K \in \mathcal{T}$ the spaces

$$X(K) = \operatorname{span}\{\psi_K \mathbf{v}\,,\ \psi_E \mathbf{w}\ :\ \mathbf{v} \in R_{k_{\mathcal{T}}}(K)^n\,,\ \mathbf{w} \in R_{k_{\mathcal{E}}}(E)^n\,, \\ E \in \mathcal{E}_K\},$$
$$Y(K) = \operatorname{span}\{\psi_K q\ :\ q \in R_{k_{\mathbf{u}}-1}(K)\},$$

where ψ_K and ψ_E are the element and edge respectively face bubble functions defined in §I.2.12 (p. 27). One can prove that the definition of $k_{\mathcal{T}}$ ensures that the local discrete problem

Find $\mathbf{u}_K \in X(K)$ and $p_K \in Y(K)$ such that

$$\int_K \nabla \mathbf{u}_K : \nabla \mathbf{v}_K \mathrm{d}x$$
$$-\int_K p_K \operatorname{div} \mathbf{v}_K \mathrm{d}x = \int_K \{\mathbf{f} + \Delta \mathbf{u}_{\mathcal{T}} - \nabla p_{\mathcal{T}}\} \cdot \mathbf{v}_K \mathrm{d}x + \int_{\partial K} \mathbb{J}_{\partial K}(\mathbf{n}_K \cdot (\nabla \mathbf{u}_{\mathcal{T}} - p_{\mathcal{T}} \underline{\mathbf{I}})) \cdot \mathbf{v}_K \mathrm{d}S$$
$$\int_K q_K \operatorname{div} \mathbf{u}_K \mathrm{d}x = \int_K q_K \operatorname{div} \mathbf{u}_{\mathcal{T}} \mathrm{d}x$$

for all $\mathbf{v}_K \in X(K)$ and all $q_K \in Y(K)$.

has a unique solution. With this solution we define the Neumann estimator by

$$\eta_{N,K} = \left\{ |\mathbf{u}_K|^2_{H^1(K)} + \|p_K\|^2_{L^2(K)} \right\}^{1/2}.$$

The above local problem is a discrete version of the Stokes equations with Neumann boundary conditions

$$\begin{aligned} -\Delta \mathbf{v} + \operatorname{grad} q &= \widetilde{\mathbf{f}} \quad \text{in } K \\ \operatorname{div} \mathbf{v} &= g \quad \text{in } K \\ \mathbf{n}_K \cdot \nabla \mathbf{v} - q\mathbf{n}_K &= \mathbf{b} \quad \text{on } \partial K \end{aligned}$$

where the data $\widetilde{\mathbf{f}}$, g, and $\mathbf{b}$ are determined by the exterior force $\mathbf{f}$ and the discrete solution $\mathbf{u}_\mathcal{T}$, $p_\mathcal{T}$.

For the Dirichlet-type estimator we denote, as in the previous subsection, for every element $K \in \mathcal{T}$ by ω_K the union of all elements in $\mathcal{T}$ that share an edge, if $n = 2$, or a face, if $n = 3$, (cf. Figure II.7.1 (p. 65)) and set

$$\begin{aligned} \widetilde{X}(K) = \operatorname{span}\{&\psi_{K'}\mathbf{v}\,,\ \psi_E\mathbf{w}\ :\ \mathbf{v} \in R_{k_\mathcal{T}}(K')^n\,,\ K' \in \mathcal{T} \cap \omega_K\,, \\ &\mathbf{w} \in R_{k_\mathcal{E}}(E)^n\,,\ E \in \mathcal{E}_K\}, \\ \widetilde{Y}(K) = \operatorname{span}\{&\psi_{K'}q\ :\ q \in R_{k_\mathbf{u}-1}(K')\,,\ K' \in \mathcal{T} \cap \omega_K\}. \end{aligned}$$

Again, one can prove that the definition of $k_\mathcal{T}$ ensures that the local discrete problem

Find $\widetilde{\mathbf{u}}_K \in \widetilde{X}(K)$ and $\widetilde{p}_K \in \widetilde{Y}(K)$ such that

$$\begin{aligned} &\int_{\omega_K} \nabla\widetilde{\mathbf{u}}_K : \nabla\mathbf{v}_K \mathrm{d}x \\ &-\int_{\omega_K} \widetilde{p}_K \operatorname{div}\mathbf{v}_K \mathrm{d}x = \int_{\omega_K} \mathbf{f}\cdot\mathbf{v}_K \mathrm{d}x - \int_{\omega_K} \nabla\mathbf{u}_\mathcal{T} : \nabla\mathbf{v}_K \mathrm{d}x \\ &\qquad\qquad + \int_{\omega_K} p_\mathcal{T} \operatorname{div}\mathbf{v}_K \mathrm{d}x \\ &\int_{\omega_K} q_K \operatorname{div}\widetilde{\mathbf{u}}_K \mathrm{d}x = \int_{\omega_K} q_K \operatorname{div}\mathbf{u}_\mathcal{T} \mathrm{d}x \end{aligned}$$

for all $\mathbf{v}_K \in \widetilde{X}(K)$ and all $q_K \in \widetilde{Y}(K)$.

has a unique solution. With this solution we define the Dirichlet estimator by

$$\eta_{D,K} = \left\{ |\widetilde{\mathbf{u}}_K|^2_{H^1(\omega_K)} + \|\widetilde{p}_K\|^2_{L^2(\omega_K)} \right\}^{1/2}.$$

This local problem is a discrete version of the Stokes equations with Dirichlet boundary conditions

$$\begin{aligned} -\Delta \mathbf{v} + \operatorname{grad} q &= \widetilde{\mathbf{f}} \quad \text{in } \omega_K \\ \operatorname{div} \mathbf{v} &= g \quad \text{in } \omega_K \\ \mathbf{v} &= 0 \quad \text{on } \partial\omega_K \end{aligned}$$

where the data $\widetilde{\mathbf{f}}$ and g are determined by the exterior force $\mathbf{f}$ and the discrete solution $\mathbf{u}_\mathcal{T}$, $p_\mathcal{T}$.

One can prove that both estimators are reliable and efficient and equivalent to the residual estimator:

$$\begin{aligned} \|\mathbf{u}-\mathbf{u}_\mathcal{T}\|_1 + \|p-p_\mathcal{T}\|_0 &\le c_1 \left\{ \sum_{K\in\mathcal{T}} \eta_{N,K}^2 \right\}^{1/2}, \\ \eta_{N,K} &\le c_2 \Big\{ \|\mathbf{u}-\mathbf{u}_\mathcal{T}\|_{H^1(\omega_K)} + \|p-p_\mathcal{T}\|_{L^2(\omega_K)} \\ &\qquad + h_K \|\mathbf{f}-\mathbf{f}_\mathcal{T}\|_{L^2(\omega_K)} \Big\}, \\ \eta_K &\le c_3 \eta_{N,K}, \\ \eta_{N,K} &\le c_4 \left\{ \sum_{K'\subset\omega_K} \eta_{K'}^2 \right\}^{1/2}, \\ \|\mathbf{u}-\mathbf{u}_\mathcal{T}\|_1 + \|p-p_\mathcal{T}\|_0 &\le c_5 \left\{ \sum_{K\in\mathcal{T}} \eta_{D,K}^2 \right\}^{1/2}, \\ \eta_{D,K} &\le c_6 \Big\{ \|\mathbf{u}-\mathbf{u}_\mathcal{T}\|_{H^1(\widetilde{\omega}_K)} + \|p-p_\mathcal{T}\|_{L^2(\widetilde{\omega}_K)} \\ &\qquad + h_K \|\mathbf{f}-\mathbf{f}_\mathcal{T}\|_{L^2(\widetilde{\omega}_K)} \Big\}, \\ \eta_K &\le c_7 \eta_{D,K}, \\ \eta_{D,K} &\le c_8 \left\{ \sum_{K'\subset\widetilde{\omega}_K} \eta_{K'}^2 \right\}^{1/2}. \end{aligned}$$

Here, $\widetilde{\omega}_K$ is the union of all elements in $\mathcal{T}$ that share at least a vertex with K (cf. Figure I.2.5 (p. 27)). The constants $c_1, \dots, c_8$ only depend on the shape parameter of the partition $\mathcal{T}$, the polynomial degree of the discretization, and the stability parameter c_Ω of the Stokes equations.

REMARK II.7.4. The computation of the estimators $\eta_{N,K}$ and $\eta_{D,K}$ obviously is more expensive than the one of the residual estimator η_K. This is recompensed by a higher accuracy. The residual estimator, on the other hand, is completely satisfactory for identifying the regions for mesh-refinement. Thus it is recommended to use the cheaper

residual estimator for refining the mesh and one of the more expensive Neumann-type or Dirichlet-type estimators for determining the actual error on the final mesh.

II.7.5. Marking strategies. There are two popular marking strategies for determining the set $\widetilde{\mathcal{T}}_k$ in the general adaptive algorithm: the *maximum strategy* II.7.2 and the *equilibration strategy* II.7.3.

Algorithm II.7.2 Maximum strategy

Require: partition $\mathcal{T}$, error estimates $(\eta_K)_{K\in\mathcal{T}}$, threshold $\theta \in (0,1)$.
Provide: subset $\widetilde{\mathcal{T}}$ of *marked* elements that should be refined.

$\widetilde{\mathcal{T}} \leftarrow \emptyset$
$\eta \leftarrow \max_{K\in\mathcal{T}} \eta_K$
for $K \in \mathcal{T}$ **do**
  if $\eta_K \geq \theta\eta$ **then**
    $\widetilde{\mathcal{T}} \leftarrow \widetilde{\mathcal{T}} \cup \{K\}$
  end if
end for

Algorithm II.7.3 Equilibration strategy

Require: partition $\mathcal{T}$, error estimates $(\eta_K)_{K\in\mathcal{T}}$, threshold $\theta \in (0,1)$.
Provide: subset $\widetilde{\mathcal{T}}$ of *marked* elements that should be refined.

$\widetilde{\mathcal{T}} \leftarrow \emptyset$, $\Sigma \leftarrow 0$, $\Theta \leftarrow \sum_{K\in\mathcal{T}} \eta_K^2$
while $\Sigma < \theta\Theta$ **do**
  $\eta \leftarrow \max_{K\in\mathcal{T}\setminus\widetilde{\mathcal{T}}} \eta_K$
  for $K \in \mathcal{T} \setminus \widetilde{\mathcal{T}}$ **do**
    if $\eta_K = \eta$ **then**
      $\widetilde{\mathcal{T}} \leftarrow \widetilde{\mathcal{T}} \cup \{K\}$, $\Sigma \leftarrow \Sigma + \eta_K^2$
    end if
  end for
end while

At the end of this algorithm II.7.3 the set $\widetilde{\mathcal{T}}$ satisfies

$$\sum_{K\in\widetilde{\mathcal{T}}} \eta_K^2 \geq \theta \sum_{K\in\mathcal{T}} \eta_K^2.$$

Both marking strategies yield comparable results. The maximum strategy obviously is cheaper than the equilibration strategy. In the maximum strategy, a large value of θ leads to small sets $\widetilde{\mathcal{T}}$, i.e. very few elements are marked and a small value of θ leads to large sets $\widetilde{\mathcal{T}}$, i.e. nearly all elements are marked. In the equilibration strategy on the contrary, a small value of θ leads to small sets $\widetilde{\mathcal{T}}$, i.e. very few elements are marked and a large value of θ leads to large sets $\widetilde{\mathcal{T}}$, i.e. nearly all elements are marked. A popular and well established choice is $\theta \approx 0.5$.

II.7.6. Regular refinement. Elements that are marked for refinement usually are refined by connecting their midpoints of edges. The resulting elements are called *red*. The corresponding refinement is called *regular*.

Triangles and quadrilaterals are thus subdivided into four smaller triangles and quadrilaterals that are similar to the parent element, i.e. have the same angles. Thus the shape parameter of the elements does not change.

This is illustrated by the top-left triangle of Figure II.6.2 (p. 60) and by the top square of Figure II.6.3 (p. 61). The numbers outside the elements indicate the local enumeration of edges and vertices of the parent element. The numbers inside the elements close to the vertices indicate the local enumeration of the vertices of the child elements. The numbers +0, +1 etc. inside the elements give the enumeration of the children. Note that the enumeration of new elements and new vertices is chosen in such a way that triangles and quadrilaterals may be treated simultaneously with a minimum of case selections.

Parallelepipeds are also subdivided into eight smaller similar parallelepipeds by joining the midpoints of edges.

For tetrahedrons, the situation is more complicated. Joining the midpoints of edges introduces four smaller similar tetrahedrons at the vertices of the parent tetrahedron plus a double pyramid in its interior. The latter one is subdivided into four small tetrahedrons by cutting it along two orthogonal planes. These tetrahedrons, however, are not similar to the parent tetrahedron. Yet there are rules which determine the cutting planes such that a repeated refinement according to these rules leads to at most four similarity classes of elements originating from a parent element. Thus these rules guarantee that the shape parameter of the partition does not deteriorate during a repeated adaptive refinement procedure.

II.7.7. Additional refinement. Since not all elements are refined regularly, we need additional refinement rules in order to avoid *hanging nodes* (cf. Figure II.7.2) and to ensure the admissibility of the refined partition. These rules are illustrated in Figures II.6.2 (p. 60) and II.6.3 (p. 61).

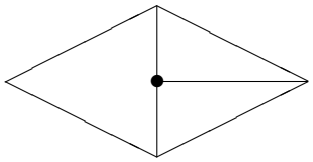

FIGURE II.7.2. Example of a hanging node

For abbreviation we call the resulting elements *green*, *blue*, and *purple*. They are obtained as follows:

- a green element by bisecting exactly one edge,
- a blue element by bisecting exactly two edges,
- a purple quadrilateral by bisecting exactly three edges.

In order to avoid too acute or too abstuse triangles, the blue and green refinement of triangles obey to the following two rules:

- In a blue refinement of a triangle, the longest one of the refinement edges is bisected first.
- Before performing a green refinement of a triangle it is checked whether the refinement edge is part of an edge which has been bisected during the last `ng` generations. If this is the case, a blue refinement is performed instead.

The second rule is illustrated in Figure II.7.3. The cross in the left part represents a hanging node which should be eliminated by a green refinement. The right part shows the blue refinement which is performed instead. Here the cross represents the new hanging node which is created by the blue refinement. Numerical experiments indicate that the optimal value of `ng` is 1. Larger values result in an excessive blow-up of the refinement zone.

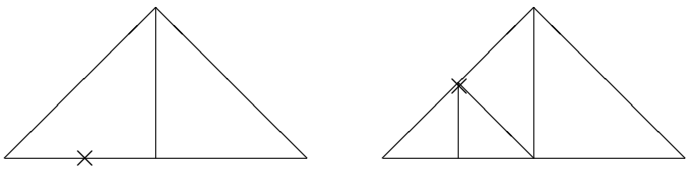

FIGURE II.7.3. Forbidden green refinement and substituting blue refinement

REMARK II.7.5. The *marked edge bisection* is an alternative to the described red-green-blue-refinement (see [11, §III.1.4]). For time-dependent problems with moving fronts, mesh refinement should be accompanied by *mesh coarsening* as described in §IV.2.4 (p. 106) below and [11, §III.1.5]. Finally, both mesh refinement and coarsening may be accompanied by *mesh smoothing* which retains the number of elements and vertices and their connectivity but changes the location of the vertices in order to improve a suitable quality measure for the mesh such as e.g. the shape parameter (see [11, §III.1.6]).

II.7.8. Required data structures. In this sub-section we shortly describe the required data structures for a `Java`, `C++`, or `Python` implementation of an adaptive finite element algorithm. For simplicity we consider only the two-dimensional case. Note that the data structures are independent of the particular differential equation and apply to all engineering problems which require the approximate solution of partial differential equations.

The class `NODE` realizes the concept of a node, i.e. of a vertex of a grid. It has three members `c`, `t`, and `d`.
The member `c` stores the co-ordinates in Euclidean 2-space. It is a double array of length 2.
The member `t` stores the type of the node. It equals 0 if it is an interior point of the computational domain. It is k, $k > 0$, if the node belongs to the k-th component of the Dirichlet boundary part of the computational domain. It equals $-k$, $k > 0$, if the node is on the k-th component of the Neumann boundary.
The member `d` gives the address of the corresponding degree of freedom. It equals -1 if the corresponding node is not a degree of freedom, e.g. since it lies on the Dirichlet boundary. This member takes into account that not every node actually is a degree of freedom.

The class `ELEMENT` realizes the concept of an element. Its member `nv` determines the element type, i.e. triangle or quadrilateral. Its members `v` and `e` realize the vertex and edge informations, respectively. Both are integer arrays of length 4.
It is assumed that `v`$[3] = -1$ if `nv`$= 3$.
A value `e`$[i] = -1$ indicates that the corresponding edge is on a straight part of the boundary. Similarly `e`$[i] = -k - 2$, $k \geq 0$, indicates that the endpoints of the corresponding edge are on the k-th curved part of the boundary. A value `e`$[i] = j \geq 0$ indicates that edge i of the current element is adjacent to element number j. Thus the member `e` decribes the neighbourhood relation of elements.
The members `p`, `c`, and `t` realize the grid hierarchy and give the number of the parent, the number of the first child, and the refinement type, respectively. In particular we have

$$\texttt{t} \in \begin{cases} \{0\} & \text{if the element is not refined} \\ \{1,\ldots,4\} & \text{if the element is refined green} \\ \{5\} & \text{if the element is refined red} \\ \{6,\ldots,24\} & \text{if the element is refined blue} \\ \{25,\ldots,100\} & \text{if the element is refined purple.} \end{cases}$$

At first sight it may seem strange to keep the information about nodes and elements in different classes. Yet this approach has several advantages:

- It minimizes the storage requirement. The co-ordinates of a node must be stored only once. If nodes and elements are represented by a common structure, these co-ordinates are stored $4 - 6$ times.
- The elements represent the topology of the grid which is independent of the particular position of the nodes. If nodes and

elements are represented by different structures it is much easier to implement mesh smoothing algorithms which affect the position of the nodes but do not change the mesh topology.

When creating a hierarchy of adaptively refined grids, the nodes are completely hierarchical, i.e. a node of grid $\mathcal{T}_i$ is also a node of any grid $\mathcal{T}_j$ with $j > i$. Since in general the grids are only partly refined, the elements are not completely hierarchical. Therefore, all elements of all grids are stored.

The information about the different grids is implemented by the class `LEVEL`. Its members `nn`, `nt`, `nq`, and `ne` give the number of nodes, triangles, quadrilaterals, and edges, resp. of a given grid. The members `first` and `last` give the addresses of the first element of the current grid and of the first element of the next grid, respectively. The member `dof` yields the number of degrees of freedom of the corresponding discrete finite element problems.

CHAPTER III

Stationary nonlinear problems

III.1. Discretization of the stationary Navier-Stokes equations

III.1.1. Variational formulation. We recall the stationary incompressible Navier-Stokes equations of §I.1.13 (p. 16) with no-slip boundary condition and the re-scaling of Remark I.1.4 (p. 15)

$$\begin{aligned} -\Delta \mathbf{u} + Re(\mathbf{u}\cdot\nabla)\mathbf{u} + \operatorname{grad} p &= \mathbf{f} \quad \text{in } \Omega \\ \operatorname{div} \mathbf{u} &= 0 \quad \text{in } \Omega \\ \mathbf{u} &= 0 \quad \text{on } \Gamma \end{aligned}$$

where $Re > 0$ is the Reynolds number. The variational formulation of this problem is given by

Find $\mathbf{u} \in H_0^1(\Omega)^n$ and $p \in L_0^2(\Omega)$ such that

$$\begin{aligned} \int_\Omega \nabla\mathbf{u} : \nabla\mathbf{v}\mathrm{d}x - \int_\Omega p \operatorname{div}\mathbf{v}\mathrm{d}x \\ + \int_\Omega Re[(\mathbf{u}\cdot\nabla)\mathbf{u}]\cdot\mathbf{v}\mathrm{d}x = \int_\Omega \mathbf{f}\cdot\mathbf{v}\mathrm{d}x \\ \int_\Omega q \operatorname{div}\mathbf{u}\mathrm{d}x = 0 \end{aligned}$$

for all $\mathbf{v} \in H_0^1(\Omega)^n$ and all $q \in L_0^2(\Omega)$.

It has the following properties:

- Any solution $\mathbf{u}$, p of the Navier-Stokes equations is a solution of the above variational problem.
- Any solution $\mathbf{u}$, p of the variational problem, which is sufficiently smooth, is a solution of the Navier-Stokes equations.

III.1.2. Fixed-point formulation. For the mathematical analysis it is convenient to write the variational formulation of the Navier-Stokes equations as a fixed-point equation. To this end we denote by T the *Stokes operator* which associates with each $\mathbf{g}$ the unique solution $\mathbf{v} = T\mathbf{g}$ of the Stokes equations:

Find $\mathbf{v} \in H_0^1(\Omega)^n$ and $q \in L_0^2(\Omega)$ such that

$$\int_\Omega \nabla \mathbf{v} : \nabla \mathbf{w} \mathrm{d}x - \int_\Omega q \operatorname{div} \mathbf{w} \mathrm{d}x = \int_\Omega \mathbf{g} \cdot \mathbf{w} \mathrm{d}x$$
$$\int_\Omega r \operatorname{div} \mathbf{v} \mathrm{d}x = 0$$

for all $\mathbf{w} \in H_0^1(\Omega)^n$ and all $r \in L_0^2(\Omega)$.
Then the variational formulation of the Navier-Stokes equations takes the equivalent fixed-point form

$$\boxed{\mathbf{u} = T(\mathbf{f} - Re(\mathbf{u} \cdot \nabla)\mathbf{u}).}$$

III.1.3. Existence and uniqueness results. The following existence and uniqueness results can be proven for the variational formulation of the Navier-Stokes equations:

- The variational problem admits at least one solution.
- Every solution of the variational problem satisfies the a priori bound
$$|\mathbf{u}|_1 \leq Re \|\mathbf{f}\|_0.$$
- The variational problem admits a unique solution provided
$$\gamma Re \|\mathbf{f}\|_0 < 1,$$
where the constant γ only depends on the domain Ω and can be estimated by
$$\gamma \leq \operatorname{diam}(\Omega)^{4-\frac{n}{2}} 2^{n-\frac{1}{2}}.$$
- Every solution of the variational problem has the same regularity properties as the solution of the Stokes equations.
- The mapping, which associates with Re a solution of the variational problem, is differentiable. Its derivative with respect to Re is a continuous linear operator which is invertible with a continuous inverse for all but countably many values of Re, i.e. there are only countably many turning or bifurcation points.

III.1.4. Finite element discretization. For the finite element discretization of the Navier-Stokes equations we choose a partition $\mathcal{T}$ of the domain Ω and corresponding finite element spaces $X(\mathcal{T})$ and $Y(\mathcal{T})$ for the velocity and pressure. *These spaces have to satisfy the*

inf-sup condition of §II.2.6 *(p. 36).* We then replace in the variational problem the spaces $H_0^1(\Omega)^n$ and $L_0^2(\Omega)$ by $X(\mathcal{T})$ and $Y(\mathcal{T})$, respectively. This leads to the following discrete problem:

Find $\mathbf{u}_\mathcal{T} \in X(\mathcal{T})$ and $p_\mathcal{T} \in Y(\mathcal{T})$ such that

$$\int_\Omega \nabla\mathbf{u}_\mathcal{T} : \nabla\mathbf{v}_\mathcal{T}\mathrm{d}x - \int_\Omega p_\mathcal{T} \operatorname{div}\mathbf{v}_\mathcal{T}\mathrm{d}x + \int_\Omega Re[(\mathbf{u}_\mathcal{T}\cdot\nabla)\mathbf{u}_\mathcal{T}]\cdot\mathbf{v}_\mathcal{T}\mathrm{d}x = \int_\Omega \mathbf{f}\cdot\mathbf{v}_\mathcal{T}\mathrm{d}x$$

$$\int_\Omega q_\mathcal{T} \operatorname{div}\mathbf{u}_\mathcal{T}\mathrm{d}x = 0$$

for all $\mathbf{v}_\mathcal{T} \in X(\mathcal{T})$ and all $q_\mathcal{T} \in Y(\mathcal{T})$.

III.1.5. Fixed-point formulation of the discrete problem. The finite element discretization of the Navier-Stokes can be written in a fixed-point form similar to the variational problem. To this end we denote by $T_\mathcal{T}$ the *discrete Stokes operator* which associates with every $\mathbf{g}$ the solution $\mathbf{v}_\mathcal{T} = T_\mathcal{T}\mathbf{g}$ of the discrete Stokes problem:
Find $\mathbf{v}_\mathcal{T} \in X(\mathcal{T})$ and $q_\mathcal{T} \in Y(\mathcal{T})$ such that

$$\int_\Omega \nabla\mathbf{v}_\mathcal{T} : \nabla\mathbf{w}_\mathcal{T}\mathrm{d}x - \int_\Omega q_\mathcal{T} \operatorname{div}\mathbf{w}_\mathcal{T}\mathrm{d}x = \int_\Omega \mathbf{g}\cdot\mathbf{w}_\mathcal{T}\mathrm{d}x$$

$$\int_\Omega r_\mathcal{T} \operatorname{div}\mathbf{v}_\mathcal{T}\mathrm{d}x = 0$$

for all $\mathbf{w}_\mathcal{T} \in X(\mathcal{T})$ and all $r_\mathcal{T} \in Y(\mathcal{T})$.
The discrete Navier-Stokes problem then takes the equivalent fixed-point form

$$\mathbf{u}_\mathcal{T} = T_\mathcal{T}(\mathbf{f} - Re(\mathbf{u}_\mathcal{T}\cdot\nabla)\mathbf{u}_\mathcal{T}).$$

Note that, as for the variational problem, this equation also determines the pressure via the operator $T_\mathcal{T}$.

III.1.6. Properties of the discrete problem. The discrete problem has similar properties as the variational problem:

- The discrete problem admits at least one solution.
- Every solution of the discrete problem satisfies the a priori bound

$$|\mathbf{u}_\mathcal{T}|_1 \leq Re\|\mathbf{f}\|_0.$$

- The discrete problem admits a unique solution provided
$$\gamma Re\|\mathbf{f}\|_0 < 1,$$
where the constant γ is the same as for the variational problem.
- The mapping, which associates with Re a solution of the discrete problem, is differentiable. Its derivative with respect to Re is a continuous linear operator which is invertible with a continuous inverse for all but finitely many values of Re, i.e. there are only finitely many turning or bifurcation points.

REMARK III.1.1. The number of turning or bifurcation points of the discrete problem of course depends on the partition $\mathcal{T}$ and on the choice of the spaces $X(\mathcal{T})$ and $Y(\mathcal{T})$.

III.1.7. Symmetrization. The integration by parts formulae of §I.2.3 (p. 18) imply for all $\mathbf{u}, \mathbf{v}, \mathbf{w} \in H_0^1(\Omega)^n$ the identity

$$\int_\Omega [(\mathbf{u}\cdot\nabla)\mathbf{v}]\cdot\mathbf{w}\mathrm{d}x = -\int_\Omega [(\mathbf{u}\cdot\nabla)\mathbf{w}]\cdot\mathbf{v}\mathrm{d}x + \int_\Omega \mathbf{v}\cdot\mathbf{w}\,\mathrm{div}\,\mathbf{u}\mathrm{d}x.$$

If $\mathbf{u}$ is solenoidal, i.e. $\mathrm{div}\,\mathbf{u} = 0$, this in particular yields

$$\int_\Omega [(\mathbf{u}\cdot\nabla)\mathbf{v}]\cdot\mathbf{w}\mathrm{d}x = -\int_\Omega [(\mathbf{u}\cdot\nabla)\mathbf{w}]\cdot\mathbf{v}\mathrm{d}x$$

for all $\mathbf{v}, \mathbf{w} \in H_0^1(\Omega)^n$.

This symmetry in general is violated for the discrete problem since $\mathrm{div}\,\mathbf{u}_\mathcal{T} \neq 0$. To enforce the symmetry, one often replaces in the discrete problem the term

$$\int_\Omega Re[(\mathbf{u}_\mathcal{T}\cdot\nabla)\mathbf{u}_\mathcal{T}]\cdot\mathbf{v}_\mathcal{T}\mathrm{d}x$$

by

$$\frac{1}{2}\int_\Omega Re[(\mathbf{u}_\mathcal{T}\cdot\nabla)\mathbf{u}_\mathcal{T}]\cdot\mathbf{v}_\mathcal{T}\mathrm{d}x - \frac{1}{2}\int_\Omega Re[(\mathbf{u}_\mathcal{T}\cdot\nabla)\mathbf{v}_\mathcal{T}]\cdot\mathbf{u}_\mathcal{T}\mathrm{d}x.$$

III.1.8. A priori error estimates. As in §II.2.11 (p. 39) we denote by $k_\mathbf{u}$ and k_p the maximal polynomial degrees k and m such that

- $S_0^{k,0}(\mathcal{T})^n \subset X(\mathcal{T})$

and either

- $S^{m,-1}(\mathcal{T})\cap L_0^2(\Omega) \subset Y(\mathcal{T})$ when using a discontinuous pressure approximation or
- $S^{m,0}(\mathcal{T}) \cap L_0^2(\Omega) \subset Y(\mathcal{T})$ when using a continuous pressure approximation.

Set

$$\boxed{k = \min\{k_{\mathbf{u}} - 1, k_p\}.}$$

Further we denote by $\mathbf{u}$, p a solution of the variational formulation of the Navier-Stokes equations and by $\mathbf{u}_{\mathcal{T}}$, $p_{\mathcal{T}}$ a solution of the discrete problem.

With these notations the following a priori error estimates can be proved:

Assume that

- $\mathbf{u} \in H^{k+2}(\Omega)^n \cap H_0^1(\Omega)^n$ and $p \in H^{k+1}(\Omega) \cap L_0^2(\Omega)$,
- $hRe\|\mathbf{f}\|_0$ is sufficiently small, and
- $\mathbf{u}$ is no turning or bifurcation point of the variational problem,

then

$$|\mathbf{u} - \mathbf{u}_{\mathcal{T}}|_1 + \|p - p_{\mathcal{T}}\|_0 \leq c_1 h^{k+1} Re^2 \|\mathbf{f}\|_k.$$

If in addition Ω is convex, then

$$\|\mathbf{u} - \mathbf{u}_{\mathcal{T}}\|_0 \leq c_2 h^{k+2} Re^2 \|\mathbf{f}\|_k.$$

The constants c_1 and c_2 only depend on the domain Ω.

Remark III.1.2. (1) As for the Stokes equations, the above regularity assumptions are not realistic for practical problems.
(2) The condition "$hRe\|\mathbf{f}\|_0$ sufficiently small" can in general not be quantified. Therefore it cannot be checked for a given discretization.
(3) One cannot conclude from the computed discrete solution whether the analytical solution is a turning or bifurcation point or not.
(4) For most practical examples, the right-hand side of the above error estimates behaves like $O(hRe^2)$ or $O(h^2Re^2)$. Thus they require unrealistically small mesh-sizes for large Reynolds numbers.

These observations show that the a priori error estimates are of purely academic interest. Practical informations can only be obtained from a posteriori error estimates (cf. §III.3 (p. 89)).

III.1.9. A warning example. We consider the one-dimensional Navier-Stokes equations

$$\begin{aligned} -u'' + Re\, uu' &= 0 \qquad \text{in } I = (-1, 1) \\ u(-1) &= 1 \\ u(1) &= -1. \end{aligned}$$

Since

$$uu' = \left(\frac{1}{2}u^2\right)',$$

we conclude that

$$-u' + \frac{Re}{2}u^2 = c$$

is constant.

To determine the constant c, we integrate the above equation and obtain

$$2c = \int_{-1}^{1} c\mathrm{d}x = \int_{-1}^{1} -u' + \frac{Re}{2}u^2\mathrm{d}x = 2 + \underbrace{\frac{Re}{2}\int_{-1}^{1} u^2\mathrm{d}x}_{\geq 0}.$$

This shows that $c \geq 1$ and that we may write $c = \gamma^2$ with a different constant $\gamma \geq 1$. Hence every solution of the differential equation satisfies

$$u' = \frac{Re}{2}u^2 - \gamma^2.$$

Therefore it must be of the form

$$u(x) = \beta_{Re}\tanh(\alpha_{Re}x)$$

with suitable parameters α_{Re} and β_{Re} depending on Re.

The boundary conditions imply that

$$\beta_{Re} = -\frac{1}{\tanh(\alpha_{Re})}.$$

Hence we have

$$\boxed{u(x) = -\frac{\tanh(\alpha_{Re}x)}{\tanh(\alpha_{Re})}.}$$

Inserting this expression in the differential equation yields

$$\begin{aligned} 0 &= \left(-u' + \frac{Re}{2}u^2\right)' \\ &= \left(\frac{\alpha_{Re}}{\tanh(\alpha_{Re})} + \left[\frac{Re}{2} - \alpha_{Re}\tanh(\alpha_{Re})\right]u^2\right)' \\ &= 2\left[\frac{Re}{2} - \alpha_{Re}\tanh(\alpha_\alpha)\right]uu'. \end{aligned}$$

Since neither u nor its derivative u' vanish identically, we obtain the defining relation

$$\boxed{2\alpha_{Re}\tanh(\alpha_{Re}) = Re}$$

for the parameter α_{Re}.

Due to the monotonicity of tanh, this equation admits for every $Re > 0$ a unique solution α_{Re}. Since $\tanh(x) \approx 1$ for $x \gg 1$, we have $\alpha_{Re} \approx \frac{Re}{2}$ for $Re \gg 1$. Hence, the solution u has a sharp interior layer

at the origin for large values of Re. Figure III.1.1 depicts the solution u for $Re = 100$.

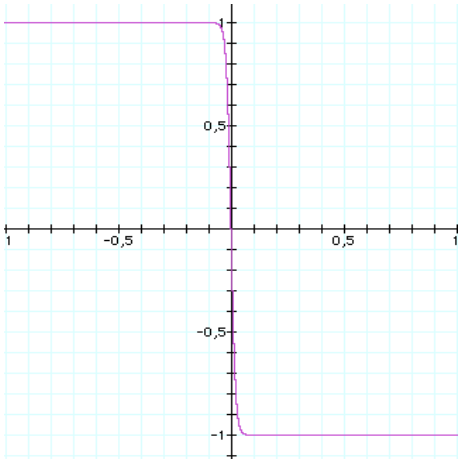

FIGURE III.1.1. Solution of the one-dimensional Navier-Stokes equations for $Re = 100$

For the discretization, we choose an integer $N \geq 1$, divide the interval $(-1, 1)$ into N small sub-intervals of equal length $h = \frac{2}{N+1}$, and use continuous piecewise linear finite elements on the resulting mesh. We denote by u_i, $0 \leq i \leq N+1$, the value of the discrete solution at the mesh-point $x_i = -1 + ih$ and evaluate all integrals with the Simpson rule. Since all integrands are piecewise polynomials of degree at most 2, this is an exact integration. With theses notations, the discretization results in the following finite difference scheme:

$$\begin{aligned} \frac{2u_i - u_{i-1} - u_{i+1}}{h} + \frac{Re}{6}(u_i - u_{i-1})(2u_i + u_{i-1}) & \\ + \frac{Re}{6}(u_{i+1} - u_i)(2u_i + u_{i+1}) = 0 \qquad & \text{for } 1 \leq i \leq N \\ u_0 = 1 & \\ u_{N+1} = -1. & \end{aligned}$$

For its solution, we try the ansatz

$$u_i = \begin{cases} 1 & \text{for } i = 0 \\ \delta & \text{for } 1 \leq i \leq N \\ -1 & \text{for } i = N+1 \end{cases}$$

with an unknown constant δ. When inserting this ansatz in the difference equations, we see that the equations corresponding to $2 \leq i \leq N-1$ are satisfied independently of the value of δ. The equations for $i = 1$ and $i = N$ on the other hand result in

$$0 = \frac{\delta - 1}{h} + \frac{Re}{6}(\delta - 1)(2\delta + 1) = \frac{\delta - 1}{h}\left[1 + \frac{Reh}{6}(2\delta + 1)\right]$$

and

$$0 = \frac{\delta+1}{h} + \frac{Re}{6}(-\delta-1)(2\delta-1) = \frac{\delta+1}{h}\left[1 - \frac{Reh}{6}(2\delta-1)\right]$$

respectively.
If $Reh = 6$, these equations have the two solutions $\delta = 1$ and $\delta = -1$. Both solutions obviously have nothing in common with the solution of the differential equation.
If $Reh \neq 6$, the above equations have no solution. Hence, our ansatz does not work. A more detailed analysis, however, shows that for $Reh \geq 6$ the discrete solution has nothing in common with the solution of the differential equation.
In summary, we see that the finite element discretization yields a qualitatively correct approximation only when $h < \frac{6}{Re}$. Hence we have to expect a prohibitively small mesh-size for real-life problems.

When using a standard symmetric finite difference approximation, things do not change. The difference equations then take the form

$$\begin{aligned} \frac{2u_i - u_{i-1} - u_{i+1}}{h} + \frac{Re}{2}u_i(u_{i+1} - u_{i-1}) = 0 \qquad & \text{for } 1 \leq i \leq N \\ u_0 = 1 & \\ u_{N+1} = -1. & \end{aligned}$$

When $Reh = 2$, our ansatz now leads to the same solution as before. Again, one can prove that one obtains a completely useless discrete solution when $Reh \geq 2$. Thus the principal result does not change in this case. Only the critical mesh-size is reduced by a factor 3.

Next, we try a backward difference approximation of the first order derivative. This results in the difference equations

$$\begin{aligned} \frac{2u_i - u_{i-1} - u_{i+1}}{h} + Reu_i(u_i - u_{i-1}) = 0 \qquad & \text{for } 1 \leq i \leq N \\ u_0 = 1 & \\ u_{N+1} = -1. & \end{aligned}$$

When trying our ansatz, we see that equations corresponding to $2 \leq i \leq N-1$ are again satisfied independently of δ. The equations for $i = 0$ and $i = N$ now take the form

$$0 = \frac{\delta-1}{h} + Re\delta(\delta-1) = \frac{\delta-1}{h}\left[1 + Reh\delta\right]$$

and

$$0 = \frac{\delta+1}{h} + Re\delta 0 = \frac{\delta+1}{h}$$

respectively. Thus, in the case $Reh = 1$ we obtain the unique solution $\delta = -1$. A more refined analysis shows, that we obtain a qualitatively correct discrete solution only if $h \leq \frac{1}{Re}$.

Finally, we try a forward difference approximation of the first order derivative. This yields the difference equations

$$\begin{aligned} \frac{2u_i - u_{i-1} - u_{i+1}}{h} + Reu_i(u_{i+1} - u_i) = 0 \qquad & \text{for } 1 \le i \le N \\ u_0 = 1 & \\ u_{N+1} = -1. & \end{aligned}$$

Our ansatz now results in the two conditions

$$0 = \frac{\delta - 1}{h} + Re\delta 0 = \frac{\delta - 1}{h}$$

and

$$0 = \frac{\delta + 1}{h} + Re\delta(-1 - \delta) = \frac{\delta + 1}{h}\left[1 - Reh\delta\right].$$

For $Reh = 1$ this yields the unique solution $\delta = 1$. A qualitatively correct discrete solution is obtained only if $h \le \frac{1}{Re}$.

These experiences show that we need a true up-winding, i.e. a backward difference approximation when $u > 0$ and a forward difference approximation when $u < 0$. Thus the up-wind direction and – with it the discretization – depend on the solution of the discrete problem!

III.1.10. Up-wind methods. The previous section shows that we must modify the finite element discretization of §III.1.4 (p. 76) in order to obtain qualitatively correct approximations also for "large" values of hRe.

One possibility to achieve this goal is to use an *up-wind difference* approximation for the convective derivative $(\mathbf{u}_\mathcal{T} \cdot \nabla)\mathbf{u}_\mathcal{T}$. To describe the idea we consider only the lowest order method.

In a first step, we approximate the integral involving the convective derivative by a one-point quadrature rule

$$\int_\Omega [(\mathbf{u}_\mathcal{T} \cdot \nabla)\mathbf{u}_\mathcal{T}] \cdot \mathbf{v}_\mathcal{T} \mathrm{d}x \approx \sum_{K \in \mathcal{T}} |K| [(\mathbf{u}_\mathcal{T}(\mathbf{x}_K) \cdot \nabla)\mathbf{u}_\mathcal{T}(\mathbf{x}_K)] \cdot \mathbf{v}_\mathcal{T}(\mathbf{x}_K).$$

Here $|K|$ is the area respectively volume of the element K and $\mathbf{x}_K$ denotes its barycentre. When using a first order approximation for the velocity, i.e. $X(\mathcal{T}) = S_0^{1,0}(\mathcal{T})^n$, this does not deteriorate the asymptotic convergence rate of the finite element discretization.

In a second step, we replace the convective derivative by a suitable up-wind difference

$$(\mathbf{u}_\mathcal{T}(\mathbf{x}_K) \cdot \nabla)\mathbf{u}_\mathcal{T}(\mathbf{x}_K) \approx \frac{1}{\|\mathbf{x}_K - \mathbf{y}_K\|} \|\mathbf{u}_\mathcal{T}(\mathbf{x}_K)\| (\mathbf{u}_\mathcal{T}(\mathbf{x}_K) - \mathbf{u}_\mathcal{T}(\mathbf{y}_K)).$$

Here $\|\cdot\|$ denotes the Euclidean norm in $\mathbb{R}^n$ and $\mathbf{y}_K$ is the intersection of the half-line $\{\mathbf{x}_K - s\mathbf{u}_\mathcal{T}(\mathbf{x}_K) \,:\, s > 0\}$ with the boundary of K (cf. Figure III.1.2).

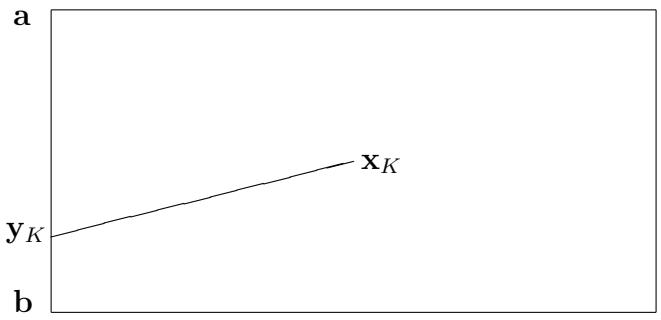

FIGURE III.1.2. Up-wind difference

In a last step, we replace $\mathbf{u}_\mathcal{T}(\mathbf{y}_K)$ by $I_\mathcal{T}\mathbf{u}_\mathcal{T}(\mathbf{y}_K)$ where $I_\mathcal{T}\mathbf{u}_\mathcal{T}$ denotes the linear interpolate of $\mathbf{u}_\mathcal{T}$ in the vertices of the edge respectively face of K which contains $\mathbf{y}_K$. If in Figure III.1.2, e.g., $\|\mathbf{y}_K - \mathbf{b}\| = \frac{1}{4}\|\mathbf{a} - \mathbf{b}\|$, we have $I_\mathcal{T}\mathbf{u}_\mathcal{T}(\mathbf{y}_K) = \frac{1}{4}\mathbf{u}_\mathcal{T}(\mathbf{a}) + \frac{3}{4}\mathbf{u}_\mathcal{T}(\mathbf{b})$.

For "large" values of hRe, this up-winding yields a better approximation than the straight-forward discretization of §III.1.4 (p. 76). For sufficiently small mesh-sizes, the error converges to zero linearly with h. The up-wind direction depends on the discrete solution. This has the awkward side-effect that the discrete nonlinear problem is not differentiable and leads to severe complications for the solution of the discrete problem.

III.1.11. The streamline-diffusion method. The *stream-line-diffusion method* has an up-wind effect, but avoids the differentiability problem of the up-wind scheme of the previous section. In particular, the resulting discrete problem is differentiable and can be solved with a Newton method. Moreover, the streamline-diffusion method simultaneously has a stabilizing effect with respect to the incompressibility constraint. In this respect it generalizes the Petrov-Galerkin method of §II.3 (p. 40).

The idea is to add an artificial viscosity in the streamline direction. Thus the solution is slightly "smeared" in its smooth direction while retaining its steep gradient in the orthogonal direction. The artificial viscosity is added via a suitable penalty term which is consistent in the sense that it vanishes for any solution of the Navier-Stokes equations.

Recalling that $\mathbb{J}_E(\cdot)$ denotes the jump across E and retaining the notations of §III.1.4 (p. 76), the *streamline-diffusion discretization* is given by:

Find $\mathbf{u}_\mathcal{T} \in X(\mathcal{T})$ and $p_\mathcal{T} \in Y(\mathcal{T})$ such that

$$\int_\Omega \nabla\mathbf{u}_\mathcal{T} : \nabla\mathbf{v}_\mathcal{T}\mathrm{d}x - \int_\Omega p_\mathcal{T} \operatorname{div}\mathbf{v}_\mathcal{T}\mathrm{d}x$$

$$
\begin{aligned}
&+\int_\Omega Re[(\mathbf{u}_\mathcal{T}\cdot\nabla)\mathbf{u}_\mathcal{T}]\cdot\mathbf{v}_\mathcal{T}\mathrm{d}x\\
&+\sum_{K\in\mathcal{T}}\delta_K h_K^2\int_K Re[-\mathbf{f}-\Delta\mathbf{u}_\mathcal{T}+\nabla p_\mathcal{T}\\
&\quad +Re(\mathbf{u}_\mathcal{T}\cdot\nabla)\mathbf{u}_\mathcal{T}]\cdot[(\mathbf{u}_\mathcal{T}\cdot\nabla)\mathbf{v}_\mathcal{T}]\mathrm{d}x\\
&\quad +\sum_{K\in\mathcal{T}}\alpha_K\delta_K\int_K \operatorname{div}\mathbf{u}_\mathcal{T}\operatorname{div}\mathbf{v}_\mathcal{T}\mathrm{d}x=\int_\Omega\mathbf{f}\cdot\mathbf{v}_\mathcal{T}\mathrm{d}x\\
&\qquad\qquad\int_\Omega q_\mathcal{T}\operatorname{div}\mathbf{u}_\mathcal{T}\mathrm{d}x\\
&\quad +\sum_{K\in\mathcal{T}}\delta_K h_K^2\int_K[-\Delta\mathbf{u}_\mathcal{T}+\nabla p_\mathcal{T}\\
&\qquad +Re(\mathbf{u}_\mathcal{T}\cdot\nabla)\mathbf{u}_\mathcal{T}]\cdot\nabla q_\mathcal{T}\mathrm{d}x\\
&\quad +\sum_{E\in\mathcal{E}}\delta_E h_E\int_E \mathbb{J}_E(p_\mathcal{T})\mathbb{J}_E(q_\mathcal{T})\mathrm{d}S=\sum_{K\in\mathcal{T}}\delta_K h_K^2\int_K\mathbf{f}\cdot\nabla q_\mathcal{T}\mathrm{d}x
\end{aligned}
$$

for all $\mathbf{v}_\mathcal{T}\in X(\mathcal{T})$ and all $q_\mathcal{T}\in Y(\mathcal{T})$.

The stabilization parameters δ_K and δ_E are chosen as described in §II.3.4 (p. 44). The additional parameter α_K has to be non-negative and can be chosen equal to zero. Computational experiments, however, suggest that the choice $\alpha_K\approx 1$ is preferable. When setting $Re=0$ and $\alpha_K=0$ the above discretization reduces to the Petrov-Galerkin discretization of the Stokes equations presented in §II.3.3 (p. 43).

III.2. Solution of the discrete nonlinear problems

III.2.1. General structure. For the solution of discrete nonlinear problems which result from a discretization of a nonlinear partial differential equation one can proceed in two ways:

- One applies a nonlinear solver, such as e.g. the Newton method, to the nonlinear differential equation and then discretizes the resulting linear partial differential equations.
- One directly applies a nonlinear solver, such as e.g. the Newton method, to the discrete nonlinear problems. The resulting discrete linear problems can then be interpreted as discretizations of suitable linear differential equations.

Both approaches often are equivalent and yield comparable approximations. In this section we will follow the first approach since it requires less notation. All algorithms can easily be re-interpreted in the second sense described above.

We recall the fixed-point formulation of the Navier-Stokes equations of §III.1.2 (p. 75)

$$\mathbf{u} = T(\mathbf{f} - Re(\mathbf{u}\cdot\nabla)\mathbf{u}) \tag{III.2.1}$$

with the Stokes operator T which associates with each $\mathbf{g}$ the unique solution $\mathbf{v} = T\mathbf{g}$ of the Stokes equations

$$\begin{aligned} -\Delta\mathbf{v} + \operatorname{grad} q &= \mathbf{g} \quad \text{in } \Omega \\ \operatorname{div}\mathbf{v} &= 0 \quad \text{in } \Omega \\ \mathbf{v} &= 0 \quad \text{on } \Gamma. \end{aligned}$$

Most of the algorithms require the solution of discrete Stokes equations or of slight variations thereof. This can be achieved with the methods of §II.6 (p. 54).

III.2.2. Fixed-point iteration. The fixed-point iteration is given by

$$\mathbf{u}^{i+1} = T(\mathbf{f} - Re(\mathbf{u}^i\cdot\nabla)\mathbf{u}^i).$$

It results in Algorithm III.2.1.

Algorithm III.2.1 Fixed-point iteration

Require: initial guess $\mathbf{u}$, tolerance $\varepsilon > 0$, maximal number of iterations N.

Provide: approximate solution of the stationary incompressible Navier-Stokes equations.

1: $D \leftarrow \infty$, $n \leftarrow 0$
2: **while** $D > \varepsilon$ and $n \leq N$ **do**
3: $\quad \mathbf{v} \leftarrow \mathbf{u}$
4: $\quad$ Solve the Stokes equations

$$\begin{aligned} -\Delta\mathbf{u} + \nabla p &= \{\mathbf{f} - Re(\mathbf{v}\cdot\nabla)\mathbf{v}\} \quad \text{in } \Omega \\ \operatorname{div}\mathbf{u} &= 0 \quad \text{in } \Omega \\ \mathbf{u} &= 0 \quad \text{on } \Gamma. \end{aligned}$$

5: $\quad D \leftarrow |\mathbf{u} - \mathbf{v}|_1$, $n \leftarrow n + 1$
6: **end while**

The fixed-point iteration converges if $Re^2\|\mathbf{f}\|_0 \leq 1$. The convergence rate approximately is $1 - Re^2\|\mathbf{f}\|_0$. Therefore this algorithm can only be recommended for problems with very small Reynolds' numbers.

III.2.3. Newton iteration. Equation (III.2.1) can be re-written in the form

$$F(\mathbf{u}) = \mathbf{u} - T(\mathbf{f} - Re(\mathbf{u}\cdot\nabla)\mathbf{u}) = 0.$$

We may apply Newton's method to F. Then we must solve in each step a linear problem of the form

$$\mathbf{g} = DF(\mathbf{u})\mathbf{v}$$

$$= \mathbf{v} + ReT((\mathbf{u}\cdot\nabla)\mathbf{v} + (\mathbf{v}\cdot\nabla)\mathbf{u}).$$

This results in Algorithm III.2.2.

Algorithm III.2.2 Newton iteration

Require: initial guess $\mathbf{u}$, tolerance $\varepsilon > 0$, maximal number of iterations N.

Provide: approximate solution of the stationary incompressible Navier-Stokes equations.

1: $D \leftarrow \infty$, $n \leftarrow 0$
2: **while** $D > \varepsilon$ and $n \leq N$ **do**
3: $\quad \mathbf{v} \leftarrow \mathbf{u}$
4: $\quad$ Solve the modified Stokes equations

$$\begin{aligned} -\Delta\mathbf{u} + \nabla p + Re(\mathbf{v}\cdot\nabla)\mathbf{u} & \\ +Re(\mathbf{u}\cdot\nabla)\mathbf{v} = \{\mathbf{f} + Re(\mathbf{v}\cdot\nabla)\mathbf{v}\} & \quad \text{in } \Omega \\ \operatorname{div}\mathbf{u} = 0 & \quad \text{in } \Omega \\ \mathbf{u} = 0 & \quad \text{on } \Gamma. \end{aligned}$$

5: $\quad D \leftarrow |\mathbf{u} - \mathbf{v}|_1$, $n \leftarrow n + 1$
6: **end while**

The Newton iteration converges quadratically. Yet, the initial guess must be "close" to the sought solution, otherwise the iteration may diverge. To avoid this, one can use a damped Newton iteration. The modified Stokes problems incorporate convection and reaction terms which are proportional to the Reynolds' number. For large Reynolds' numbers this may cause severe difficulties with the solution of the modified Stokes problems.

III.2.4. Path tracking. Instead of the single problem (III.2.1) we now look at a whole family of Navier-Stokes equations with a parameter λ

$$\mathbf{u}_\lambda = T(\mathbf{f} - \lambda(\mathbf{u}_\lambda\cdot\nabla)\mathbf{u}_\lambda).$$

Its solutions $\mathbf{u}_\lambda$ depend differentiably on the parameter λ. The derivative $\mathbf{v}_\lambda = \frac{d\mathbf{u}_\lambda}{d\lambda}$ solves the modified Stokes equations

$$\mathbf{v}_\lambda = -T(\lambda(\mathbf{v}_\lambda\cdot\nabla)\mathbf{u}_\lambda + \lambda(\mathbf{u}_\lambda\cdot\nabla\mathbf{v}_\lambda) + (\mathbf{u}_\lambda\cdot\nabla)\mathbf{u}_\lambda).$$

If we know the solution $\mathbf{u}_{\lambda_0}$ corresponding to the parameter λ_0, we can compute $\mathbf{v}_{\lambda_0}$ and may use $\mathbf{u}_{\lambda_0} + (\lambda_1 - \lambda_0)\mathbf{v}_{\lambda_0}$ as initial guess for the Newton iteration applied to the problem with parameter $\lambda_1 > \lambda_0$. If $\lambda_1 - \lambda_0$ is not too large, a few Newton iterations will yield a sufficiently good approximation of $\mathbf{v}_{\lambda_1}$.

This idea leads to Algorithm III.2.3. It should be combined with a step-length control: If the Newton algorithm in step 3 does not converge sufficiently well, the increment $\Delta\lambda$ should be reduced.

Algorithm III.2.3 Path tracking

Require: Reynolds' number Re, initial parameter $0 \leq \lambda < Re$, increment $\Delta\lambda > 0$, tolerance $\varepsilon > 0$.

Provide: approximate solution of the stationary incompressible Navier-Stokes equations with Reynolds' number Re.

1: $\mathbf{u}_\lambda \leftarrow 0$

2: **while** $\lambda < Re$ **do**

3: Apply a few Newton iterations to the Navier-Stokes equations with Reynolds' number λ, initial guess $\mathbf{u}_\lambda$, and tolerance ε. Denote the result by $\mathbf{u}_\lambda$.

4: Solve the modified Stokes equations

$$\begin{aligned} -\Delta\mathbf{v}_\lambda + \nabla q_\lambda + \lambda(\mathbf{u}_\lambda \cdot \nabla)\mathbf{v}_\lambda & \\ +\lambda(\mathbf{v}_\lambda \cdot \nabla)\mathbf{u}_\lambda &= \{\mathbf{f} - \lambda(\mathbf{u}_\lambda \cdot \nabla)\mathbf{u}_\lambda\} \quad \text{in } \Omega \\ \operatorname{div} \mathbf{v}_\lambda &= 0 \quad \text{in } \Omega \\ \mathbf{v}_\lambda &= 0 \quad \text{on } \Gamma. \end{aligned}$$

5: $\mathbf{u}_\lambda \leftarrow \mathbf{u}_\lambda + \Delta\lambda\mathbf{v}_\lambda$, $\lambda \leftarrow \min\{Re, \lambda + \Delta\lambda\}$

6: **end while**

III.2.5. Operator splitting. The idea is to decouple the difficulties associated with the nonlinear convection term and with the incompressibility constraint. This leads to Algorithm III.2.4. The nonlinear problem in step 4 is solved with one of the algorithms presented in the previous sub-sections. This task is simplified by the fact that the incompressibility condition and the pressure are now missing.

III.2.6. A nonlinear CG-algorithm. The idea is to apply a nonlinear CG-algorithm to the least-squares minimization problem

$$\text{minimize } \frac{1}{2}|\mathbf{u} - T(\mathbf{f} - Re(\mathbf{u} \cdot \nabla)\mathbf{u})|_1^2.$$

It leads to Algorithm III.2.5 (p. 93). It has the advantage that it only requires the solution of Stokes problems and thus avoids the difficulties associated with large convection and reaction terms.

III.2.7. Multigrid algorithms. When applying the multigrid algorithm of §II.6.5 (p. 57) to nonlinear problems one only has to modify the pre- and post-smoothing steps. This can be done in two possible ways:

- Apply a few iterations of the Newton algorithm to the nonlinear problem combined with very few iterations of a classical iterative scheme, such as e.g. Gauß-Seidel iteration, for the auxiliary linear problems that must be solved during the Newton iteration.

Algorithm III.2.4 Operator splitting

Require: initial guess $\mathbf{u}$, damping parameter $\omega \in (0,1)$, tolerance $\varepsilon > 0$, maximal number of iterations N.

Provide: approximate solution of the stationary incompressible Navier-Stokes equations.

1: $D \leftarrow \infty$, $n \leftarrow 0$

2: **while** $D > \varepsilon$ and $n \leq N$ **do**

3: Solve the Stokes equations

$$\begin{aligned} 2\omega\mathbf{v} - \Delta\mathbf{v} + \nabla q &= 2\omega\mathbf{u} + \mathbf{f} - Re(\mathbf{u}\cdot\nabla)\mathbf{u} && \text{in } \Omega \\ \operatorname{div}\mathbf{v} &= 0 && \text{in } \Omega \\ \mathbf{v} &= 0 && \text{on } \Gamma. \end{aligned}$$

4: Solve the nonlinear Poisson equation

$$\begin{aligned} \omega\mathbf{w} - \Delta\mathbf{w} + Re(\mathbf{w}\cdot\nabla)\mathbf{w} &= \omega\mathbf{v} + \mathbf{f} - \nabla q && \text{in } \Omega \\ \mathbf{w} &= 0 && \text{on } \Gamma. \end{aligned}$$

5: Solve the Stokes equations

$$\begin{aligned} 2\omega\mathbf{z} - \Delta\mathbf{z} + \nabla r &= 2\omega\mathbf{w} + \mathbf{f} - Re(\mathbf{w}\cdot\nabla)\mathbf{w} && \text{in } \Omega \\ \operatorname{div}\mathbf{z} &= 0 && \text{in } \Omega \\ \mathbf{z} &= 0 && \text{on } \Gamma. \end{aligned}$$

6: $D \leftarrow |\mathbf{u} - \mathbf{z}|_1$, $\mathbf{u} \leftarrow \mathbf{z}$, $p \leftarrow r$, $n \leftarrow n + 1$

7: **end while**

- Successively choose a node and the corresponding equation, freeze all unknowns that do not belong to the current node, and solve for the current unknown by applying a few Newton iterations to the resulting nonlinear equation in one unknown.

The second variant is often called *nonlinear Gauß-Seidel algorithm.*

Alternatively one can apply a variant of the multigrid algorithm of §II.6.5 (p. 57) to the linear problems that must be solved in the algorithms described above. Due to the lower costs for implementation, this variant is often preferred in practice.

III.3. Adaptivity for nonlinear problems

III.3.1. General structure. The general structure of an adaptive algorithm as described in §II.7.2 (p. 63) directly applies to nonlinear problems. One only has to adapt the a posteriori error estimator. The marking strategy, the regular refinement, the additional refinement, and the data structures described in §II.7.5 (p. 69) – §II.7.8 (p. 71) do not change.

Throughout this section we denote by $X(\mathcal{T})$ and $Y(\mathcal{T})$ finite element spaces for the velocity and pressure, respectively associated with a given partition $\mathcal{T}$ of the domain Ω. With these spaces we associate a

discretization of the stationary incompressible Navier-Stokes equations as described in §III.1 (p. 75). The computed solution of the corresponding discrete problem is denoted by $\mathbf{u}_\mathcal{T}$, $p_\mathcal{T}$, whereas $\mathbf{u}$, p denotes a solution of the Navier-Stokes equations.

III.3.2. A residual a posteriori error estimator. The residual a posteriori error estimator for the Navier-Stokes equations is given by

$$\eta_K = \Big\{ h_K^2 \|\mathbf{f} + \Delta \mathbf{u}_\mathcal{T} - \nabla p_\mathcal{T} - Re(\mathbf{u}_\mathcal{T} \cdot \nabla)\mathbf{u}_\mathcal{T}\|^2_{L^2(K)} + \|\operatorname{div} \mathbf{u}_\mathcal{T}\|^2_{L^2(K)} + \frac{1}{2} \sum_{E \in \mathcal{E}_K} h_E \|\mathbb{J}_E(\mathbf{n}_E \cdot (\nabla \mathbf{u}_\mathcal{T} - p_\mathcal{T} \underline{\mathbf{I}}))\|^2_{L^2(E)} \Big\}^{1/2}.$$

When comparing it with the residual estimator for the Stokes equations, we observe that the nonlinear convection term has been added to the element residual.

Under suitable conditions on the solution of the Navier-Stokes equations, one can prove that, as in the linear case, the residual error estimator is reliable and efficient.

III.3.3. Error estimators based on the solution of auxiliary problems. The error estimators based on the solution of auxiliary local discrete problems are very similar to those in the linear case. In particular the nonlinearity only enters into the data, the local problems itself remain linear.

We recall the notations of §II.7.4 (p. 65) and set

$$\begin{aligned}
X(K) &= \operatorname{span}\{\psi_K \mathbf{v}\,,\ \psi_E \mathbf{w}\ :\ \mathbf{v} \in R_{k_\mathcal{T}}(K)^n\,,\ \mathbf{w} \in R_{k_\mathcal{E}}(E)^n\,,\\
&\qquad\qquad E \in \mathcal{E}_K\},\\
Y(K) &= \operatorname{span}\{\psi_K q\ :\ q \in R_{k_\mathbf{u}-1}(K)\},\\
\widetilde{X}(K) &= \operatorname{span}\{\psi_{K'} \mathbf{v}\,,\ \psi_E \mathbf{w}\ :\ \mathbf{v} \in R_{k_\mathcal{T}}(K')^n\,,\ K' \in \mathcal{T} \cap \omega_K\,,\\
&\qquad\qquad \mathbf{w} \in R_{k_\mathcal{E}}(E)^n\,,\ E \in \mathcal{E}_K\},\\
\widetilde{Y}(K) &= \operatorname{span}\{\psi_{K'} q\ :\ q \in R_{k_\mathbf{u}-1}(K')\,,\ K' \in \mathcal{T} \cap \omega_K\},
\end{aligned}$$

where

$$\begin{aligned}
k_\mathcal{T} &= \max\{k_\mathbf{u}(k_\mathbf{u} - 1), k_\mathbf{u} + n, k_p - 1\},\\
k_\mathcal{E} &= \max\{k_\mathbf{u} - 1, k_p\},
\end{aligned}$$

and $k_\mathbf{u}$ and k_p denote the polynomial degrees of the velocity and pressure approximation respectively. Note that the definition of $k_\mathcal{T}$ differs

from the one in §II.7.4 (p. 65) and takes into account the polynomial degree of the nonlinear convection term.

With these notations we consider the following discrete Stokes problem with Neumann boundary conditions

Find $\mathbf{u}_K \in X(K)$ and $p_K \in Y(K)$ such that

$$\begin{aligned}
&\int_K \nabla \mathbf{u}_K : \nabla \mathbf{v}_K \mathrm{d}x \\
&-\int_K p_K \operatorname{div} \mathbf{v}_K \mathrm{d}x = \int_K \{\mathbf{f} + \Delta \mathbf{u}_\mathcal{T} - \nabla p_\mathcal{T} \\
&\qquad\qquad - Re(\mathbf{u}_\mathcal{T} \cdot \nabla)\mathbf{u}_\mathcal{T}\} \cdot \mathbf{v}_K \mathrm{d}x \\
&\qquad\qquad + \int_{\partial K} \mathbb{J}_{\partial K}(\mathbf{n}_K \cdot (\nabla \mathbf{u}_\mathcal{T} - p_\mathcal{T} \underline{\mathbf{I}})) \cdot \mathbf{v}_K \mathrm{d}S \\
&\int_K q_K \operatorname{div} \mathbf{u}_K \mathrm{d}x = \int_K q_K \operatorname{div} \mathbf{u}_\mathcal{T} \mathrm{d}x
\end{aligned}$$

for all $\mathbf{v}_K \in X(K)$ and all $q_K \in Y(K)$.

With the solution of this problem, we define the Neumann estimator by

$$\eta_{N,K} = \left\{ |\mathbf{u}_K|^2_{H^1(K)} + \|p_K\|^2_{L^2(K)} \right\}^{1/2}.$$

Similarly we can consider the following discrete Stokes problem with Dirichlet boundary conditions

Find $\widetilde{\mathbf{u}}_K \in \widetilde{X}(K)$ and $\widetilde{p}_K \in \widetilde{Y}(K)$ such that

$$\begin{aligned}
&\int_{\omega_K} \nabla \widetilde{\mathbf{u}}_K : \nabla \mathbf{v}_K \mathrm{d}x \\
&-\int_{\omega_K} \widetilde{p}_K \operatorname{div} \mathbf{v}_K \mathrm{d}x = \int_{\omega_K} \mathbf{f} \cdot \mathbf{v}_K \mathrm{d}x - \int_{\omega_K} \nabla \mathbf{u}_\mathcal{T} : \nabla \mathbf{v}_K \mathrm{d}x \\
&\qquad\qquad + \int_{\omega_K} p_\mathcal{T} \operatorname{div} \mathbf{v}_K \mathrm{d}x \\
&\qquad\qquad - \int_{\omega_K} Re(\mathbf{u}_\mathcal{T} \cdot \nabla)\mathbf{u}_\mathcal{T} \cdot \mathbf{v}_K \mathrm{d}x \\
&\int_{\omega_K} q_K \operatorname{div} \widetilde{\mathbf{u}}_K \mathrm{d}x = \int_{\omega_K} q_K \operatorname{div} \mathbf{u}_\mathcal{T} \mathrm{d}x
\end{aligned}$$

for all $\mathbf{v}_K \in \widetilde{X}(K)$ and all $q_K \in \widetilde{Y}(K)$.

With the solution of this problem, we define the Dirichlet estimator by

$$\eta_{D,K} = \left\{ |\widetilde{\mathbf{u}}_K|^2_{H^1(\omega_K)} + \|\widetilde{p}_K\|^2_{L^2(\omega_K)} \right\}^{1/2}.$$

As in the linear case, one can prove that both estimators are reliable and efficient and comparable to the residual estimator. Remark II.7.4 (p. 68) also applies to the nonlinear problem.

Algorithm III.2.5 Non-linear CG-algorithm of Polak-Ribière

Require: initial guess $\mathbf{u}$, tolerance $\varepsilon > 0$, maximal number of iterations N.

Provide: approximate solution of the stationary incompressible Navier-Stokes equations.

1: Compute the solution $\mathbf{z}$ the Stokes problem

$$\begin{aligned} -\Delta\mathbf{z} + \nabla r &= Re(\mathbf{u}\cdot\nabla)\mathbf{u} - \mathbf{f} && \text{in } \Omega \\ \operatorname{div}\mathbf{z} &= 0 && \text{in } \Omega \\ \mathbf{z} &= 0 && \text{on } \Gamma. \end{aligned}$$

2: Compute the solution $\widetilde{\mathbf{g}}$ the Stokes problem

$$\begin{aligned} -\Delta\widetilde{\mathbf{g}} + \nabla\widetilde{s} &= Re\{(\mathbf{u}+\mathbf{z})\cdot(\nabla\mathbf{u}) - (\mathbf{u}\cdot\nabla)(\mathbf{u}+\mathbf{z})\} && \text{in } \Omega \\ \operatorname{div}\widetilde{\mathbf{g}} &= 0 && \text{in } \Omega \\ \widetilde{\mathbf{g}} &= 0 && \text{on } \Gamma. \end{aligned}$$

3: $\mathbf{w} \leftarrow \mathbf{u}+\mathbf{z}+\widetilde{\mathbf{g}}$, $\mathbf{g} \leftarrow \mathbf{w}$, $E \leftarrow |\mathbf{w}|_1$, $n \leftarrow 0$

4: **while** $E > \varepsilon$ and $n \leq N$ **do**

5: Compute the solution $\mathbf{z}_1$of the Stokes problem

$$\begin{aligned} -\Delta\mathbf{z}_1 + \nabla r_1 &= Re\{(\mathbf{u}\cdot\nabla)\mathbf{w} + (\mathbf{w}\cdot\nabla)\mathbf{u}\} && \text{in } \Omega \\ \operatorname{div}\mathbf{z}_1 &= 0 && \text{in } \Omega \\ \mathbf{z}_1 &= 0 && \text{on } \Gamma \end{aligned}$$

6: Compute the solution $\mathbf{z}_2$ of the Stokes problem

$$\begin{aligned} -\Delta\mathbf{z}_2 + \nabla r_2 &= Re(\mathbf{w}\cdot\nabla)\mathbf{w} && \text{in } \Omega \\ \operatorname{div}\mathbf{z}_2 &= 0 && \text{in } \Omega \\ \mathbf{z}_2 &= 0 && \text{on } \Gamma. \end{aligned}$$

7: $\alpha \leftarrow -\int_\Omega \nabla(\mathbf{u}+\mathbf{z}) : \nabla(\mathbf{w}+\mathbf{z}_1)\mathrm{d}x$

8: $\beta \leftarrow \int_\Omega |\nabla(\mathbf{w}+\mathbf{z}_1)|^2\mathrm{d}x + \int_\Omega \nabla(\mathbf{u}+\mathbf{z}) : \nabla\mathbf{z}_2\mathrm{d}x$

9: $\gamma \leftarrow -\frac{3}{2}\int_\Omega \nabla(\mathbf{w}+\mathbf{z}_1) : \nabla\mathbf{z}_2\mathrm{d}x$

10: $\delta \leftarrow \frac{1}{2}\int_\Omega |\nabla\mathbf{z}_2|^2\mathrm{d}x$

11: Determine the smallest positive zero ρ of $\alpha + \beta x + \gamma x^2 + \delta x^3$.

12: $\mathbf{u} \leftarrow \mathbf{u} - \rho\mathbf{w}$, $\mathbf{z} \leftarrow \mathbf{z} - \rho\mathbf{z}_1 + \frac{1}{2}\rho^2\mathbf{z}_2$

13: Compute the solution $\widetilde{\mathbf{g}}$ of the Stokes problem

$$\begin{aligned} -\Delta\widetilde{\mathbf{g}} + \nabla\widetilde{s} &= Re\{(\mathbf{u}+\mathbf{z})\cdot(\nabla\mathbf{u}) \\ &\qquad - (\mathbf{u}\cdot\nabla)(\mathbf{u}+\mathbf{z})\} && \text{in } \Omega \\ \operatorname{div}\widetilde{\mathbf{g}} &= 0 && \text{in } \Omega \\ \widetilde{\mathbf{g}} &= 0 && \text{on } \Gamma \end{aligned}$$

14: $\widehat{\mathbf{g}} \leftarrow \mathbf{g}$, $\mathbf{g} \leftarrow \widetilde{\mathbf{g}} + \mathbf{u} + \mathbf{z}$

15: $\sigma \leftarrow \left\{\int_\Omega |\nabla\mathbf{g}|^2\mathrm{d}x\right\}^{-1} \int_\Omega \nabla(\mathbf{g}-\widehat{\mathbf{g}}) : \nabla\mathbf{g}\mathrm{d}x$

16: $\mathbf{w} \leftarrow \mathbf{g} + \sigma\mathbf{w}$, $E \leftarrow |\mathbf{w}|_1$, $n \leftarrow n+1$

17: **end while**

CHAPTER IV

Instationary problems

IV.1. Discretization of the instationary Navier-Stokes equations

IV.1.1. Variational formulation. We recall the instationary incompressible Navier-Stokes equations of §I.1.12 (p. 15) with no-slip boundary condition

$$\begin{aligned} \frac{\partial \mathbf{u}}{\partial t} - \nu\Delta\mathbf{u} + (\mathbf{u}\cdot\nabla)\mathbf{u} - \operatorname{grad} p &= \mathbf{f} && \text{in } \Omega\times(0,T)\\ \operatorname{div}\mathbf{u} &= 0 && \text{in } \Omega\times(0,T)\\ \mathbf{u} &= 0 && \text{on } \Gamma\times(0,T)\\ \mathbf{u}(\cdot,0) &= \mathbf{u}_0 && \text{in } \Omega. \end{aligned}$$

Here, $T>0$ is given final time and $\mathbf{u}_0$ denotes a given initial velocity.

The variational formulation of this problem is given by

Find a velocity field $\mathbf{u}$ with

$$\max_{0<t<T}\int_\Omega |\mathbf{u}(x,t)|^2\mathrm{d}x + \int_0^T\int_\Omega |\nabla\mathbf{u}(x,t)|^2\mathrm{d}x\mathrm{d}t < \infty$$

and a pressure p with

$$\int_0^T\int_\Omega |p(x,t)|^2\mathrm{d}x\mathrm{d}t < \infty$$

such that

$$\begin{aligned} \int_0^T\int_\Omega \Big\{ &-\mathbf{u}(x,t)\frac{\partial\mathbf{v}(x,t)}{\partial t} + \nu\nabla\mathbf{u}(x,t) : \nabla\mathbf{v}(x,t)\\ &+ [(\mathbf{u}(x,t)\cdot\nabla)\mathbf{u}(x,t)]\cdot\mathbf{v}(x,t)\\ &- p(x,t)\operatorname{div}\mathbf{v}(x,t)\Big\}\mathrm{d}x\mathrm{d}t\\ = \int_0^T\int_\Omega \mathbf{f}(x,t)\cdot\mathbf{v}(x,t)\mathrm{d}x\mathrm{d}t &+ \int_\Omega \mathbf{u}_0(x)\cdot\mathbf{v}(x,0)\mathrm{d}x \end{aligned}$$

$$\int_0^T \int_\Omega q(x,t)\,\mathrm{div}\,\mathbf{u}(x,t)\mathrm{d}x\mathrm{d}t = 0$$

holds for all $\mathbf{v}$ with

$$\max_{0<t<T} \int_\Omega \left\{ |\frac{\partial \mathbf{v}(x,t)}{\partial t}|^2 \mathrm{d}x + |\nabla \mathbf{v}(x,t)|^2 \right\} \mathrm{d}x < \infty$$

and all q with

$$\max_{0<t<T} \int_\Omega |q(x,t)|^2 \mathrm{d}x < \infty.$$

IV.1.2. Existence and uniqueness results. The variational formulation of the instationary incompressible Navier-Stokes equations admits at least one solution. In two space dimensions, this solution is unique. In three space dimensions, uniqueness of the solution can only be guaranteed within a restricted class of more regular functions. Yet, the existence of such a solution cannot be guaranteed.

Any solution of the instationary incompressible Navier-Stokes equations behaves like $\sqrt{t}$ for small times t. For larger times the smoothness with respect to time is better. This singular behaviour for small times must be taken into account for the time-discretization.

IV.1.3. Numerical methods for ordinary differential equations revisited. Before presenting the various discretization schemes for the time-dependent Navier-Stokes equations, we shortly recapitulate some basic facts on numerical methods for ordinary differential equations.

Suppose that we have to solve an initial value problem

$$\begin{aligned} \frac{d\mathbf{y}}{dt} &= \mathbf{F}(\mathbf{y},t) \\ \mathbf{y}(0) &= \mathbf{y}_0 \end{aligned} \tag{IV.1.1}$$

in $\mathbb{R}^n$ on the time interval $(0,T)$.

We choose an integer $N \geq 1$ and intermediate times $0 = t_0 < t_1 < \ldots < t_N = T$ and set $\tau_i = t_i - t_{i-1}$, $1 \leq i \leq N$. The approximation to $\mathbf{y}(t_i)$ is denoted by $\mathbf{y}^i$.

The simplest and most popular method is the *θ-scheme*. It is given by

$$\begin{aligned} \mathbf{y}^0 &= \mathbf{y}_0 \\ \frac{\mathbf{y}^i - \mathbf{y}^{i-1}}{\tau_i} &= \theta \mathbf{F}(\mathbf{y}^i, t_i) + (1-\theta)\mathbf{F}(\mathbf{y}^{i-1}, t_{i-1}), \quad 1 \leq i \leq N, \end{aligned}$$

where $\theta \in [0,1]$ is a fixed parameter. The choice $\theta = 0$ yields the *explicit Euler scheme*, $\theta = 1$ corresponds to the *implicit Euler scheme*, and the choice $\theta = \frac{1}{2}$ gives the *Crank-Nicolson scheme*. The θ-scheme is implicit unless $\theta = 0$. It is A-stable provided $\theta \geq \frac{1}{2}$. The θ-scheme is of order 1, if $\theta \neq \frac{1}{2}$, and of order 2, if $\theta = \frac{1}{2}$.

Another class of popular methods is given by the various *Runge-Kutta schemes*. They take the form

$$\begin{aligned} \mathbf{y}^0 &= \mathbf{y}_0 \\ \mathbf{y}^{i,j} &= \mathbf{y}^{i-1} + \tau_i \sum_{k=1}^{r} a_{jk} \mathbf{F}(t_{i-1} + c_k \tau_i, \mathbf{y}^{i,k}), \quad 1 \leq j \leq r, \\ \mathbf{y}^i &= \mathbf{y}^{i-1} + \tau_i \sum_{k=1}^{r} b_k \mathbf{F}(t_{i-1} + c_k \tau_i, \mathbf{y}^{i,k}), \quad 1 \leq i \leq N, \end{aligned}$$

where $0 \leq c_1 \leq \ldots \leq c_r \leq 1$. The number r is called *stage number* of the Runge-Kutta scheme. The scheme is called *explicit*, if $a_{jk} = 0$ for all $k \geq j$, otherwise it is called *implicit*.

For the ease of notation one usually collects the numbers c_k, a_{jk}, b_k in a table of the form

$$\begin{array}{c|cccc} c_1 & a_{11} & a_{12} & \ldots & a_{1r} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ c_r & a_{r1} & a_{r2} & \ldots & a_{rr} \\ \hline & b_1 & b_2 & \ldots & b_r \end{array}$$

The Euler and Crank-Nicolson schemes are Runge-Kutta schemes. They correspond to the tables

$$\begin{array}{c|c} 0 & 0 \\ \hline & 1 \end{array} \qquad \text{explicit Euler,}$$

$$\begin{array}{c|c} 1 & 1 \\ \hline & 1 \end{array} \qquad \text{implicit Euler,}$$

$$\begin{array}{c|cc} 0 & 0 & 0 \\ 1 & \frac{1}{2} & \frac{1}{2} \\ \hline & \frac{1}{2} & \frac{1}{2} \end{array} \qquad \text{Crank-Nicolson.}$$

Strongly diagonal implicit Runge-Kutta schemes (in short *SDIRK schemes*) are particularly well suited for the discretization of time-dependent partial differential equations since they combine high order with good stability properties. The simplest representatives of this class are called SDIRK 2 and SDIRK 5. They are both A-stable and

have orders 3 and 4, respectively. They are given by the data

$$\begin{array}{c|cc} \frac{3+\sqrt{3}}{6} & \frac{3+\sqrt{3}}{6} & 0 \\ \frac{3-\sqrt{3}}{6} & \frac{-\sqrt{3}}{3} & \frac{3+\sqrt{3}}{6} \\ \hline & \frac{1}{2} & \frac{1}{2} \end{array} \qquad \text{SDIRK2}$$

$$\begin{array}{c|ccccc} \frac{1}{4} & \frac{1}{4} & 0 & 0 & 0 & 0 \\ \frac{3}{4} & \frac{1}{2} & \frac{1}{4} & 0 & 0 & 0 \\ \frac{11}{20} & \frac{17}{50} & -\frac{1}{25} & \frac{1}{4} & 0 & 0 \\ \frac{1}{2} & \frac{371}{1360} & -\frac{137}{2720} & \frac{15}{544} & \frac{1}{4} & 0 \\ 1 & \frac{25}{24} & -\frac{49}{48} & \frac{125}{16} & -\frac{85}{12} & \frac{1}{4} \\ \hline & \frac{25}{24} & -\frac{49}{48} & \frac{125}{16} & -\frac{85}{12} & \frac{1}{4} \end{array} \qquad \text{SDIRK5.}$$

IV.1.4. Method of lines. This is the simplest discretization scheme for time-dependent partial differential equations.

We choose a partition $\mathcal{T}$ of the spatial domain Ω and associated finite element spaces $X(\mathcal{T})$ and $Y(\mathcal{T})$ for the velocity and pressure, respectively. *These spaces must satisfy the inf-sup condition of* §II.2.6 *(p. 36).* Then we replace in the variational formulation the velocities $\mathbf{u}$, $\mathbf{v}$ and the pressures p, q by discrete functions $\mathbf{u}_\mathcal{T}$, $\mathbf{v}_\mathcal{T}$ and $p_\mathcal{T}$, $q_\mathcal{T}$ which depend upon time and which – for every time t – have their values in $X(\mathcal{T})$ and $Y(\mathcal{T})$, respectively. Next, we choose a bases for the spaces $X(\mathcal{T})$ and $Y(\mathcal{T})$ and identify $\mathbf{u}_\mathcal{T}$ and $p_\mathcal{T}$ with their coefficient vectors with respect to the bases. Then we obtain the following *differential algebraic equation* for $\mathbf{u}_\mathcal{T}$ and $p_\mathcal{T}$

$$\begin{aligned} \frac{d\mathbf{u}_\mathcal{T}}{dt} &= \mathbf{f}_\mathcal{T} - \nu A_\mathcal{T}\mathbf{u}_\mathcal{T} - B_\mathcal{T} p_\mathcal{T} - N_\mathcal{T}(\mathbf{u}_\mathcal{T}) \\ B_\mathcal{T}^T \mathbf{u}_\mathcal{T} &= 0. \end{aligned}$$

Denoting by $\mathbf{w}_i$, $1 \le i \le \dim X(\mathcal{T})$, and r_j, $1 \le j \le \dim Y(\mathcal{T})$, the bases functions of $X(\mathcal{T})$ and $Y(\mathcal{T})$ respectively, the coefficients of the stiffness matrices $A_\mathcal{T}$ and $B_\mathcal{T}$ and of the vectors $\mathbf{f}_\mathcal{T}$ and $N_\mathcal{T}(\mathbf{u}_\mathcal{T})$ are given by

$$\begin{aligned} A_{\mathcal{T}ij} &= \int_\Omega \nabla \mathbf{w}_i(x) : \nabla \mathbf{w}_j(x) \mathrm{d}x, \\ B_{\mathcal{T}ij} &= -\int_\Omega r_j(x) \operatorname{div} \mathbf{w}_i(x) \mathrm{d}x, \\ \mathbf{f}_{\mathcal{T}i} &= \int_\Omega \mathbf{f}(x,t) \cdot \mathbf{w}_i(x) \mathrm{d}x, \end{aligned}$$

$$N_{\mathcal{T}}(\mathbf{u}_{\mathcal{T}})_i = \int_\Omega [(\mathbf{u}(x,t)\cdot\nabla)\mathbf{u}(x,t)]\cdot\mathbf{w}_i(x)\mathrm{d}x.$$

Formally, this problem can be cast in the form (IV.1.1) by working with test-functions $\mathbf{v}_{\mathcal{T}}$ in the space

$$V(\mathcal{T}) = \left\{\mathbf{v}_{\mathcal{T}} \in X(\mathcal{T}) : \int_\Omega q_{\mathcal{T}} \operatorname{div}\mathbf{v}_{\mathcal{T}}\mathrm{d}x = 0 \text{ for all } q_{\mathcal{T}} \in Y(\mathcal{T})\right\}$$

of discretely solenoidal functions. The function $\mathbf{F}$ in (IV.1.1) is then given by

$$\mathbf{F}(\mathbf{y},t) = \mathbf{f}_{\mathcal{T}} - \nu A_{\mathcal{T}}\mathbf{y} - N_{\mathcal{T}}(\mathbf{y}).$$

If we apply the θ-scheme and denote by $\mathbf{u}^i_{\mathcal{T}}$ and $p^i_{\mathcal{T}}$ the approximate values of $\mathbf{u}_{\mathcal{T}}$ and $p_{\mathcal{T}}$ at time t_i, we obtain the fully discrete scheme

$$\begin{aligned}
\mathbf{u}^0_{\mathcal{T}} &= I_{\mathcal{T}}\mathbf{u}_0\\
\frac{\mathbf{u}^i_{\mathcal{T}} - \mathbf{u}^{i-1}_{\mathcal{T}}}{\tau_i} &= -B_{\mathcal{T}}p^i_{\mathcal{T}} + \theta\left\{\mathbf{f}_{\mathcal{T}}(t_i) - \nu A_{\mathcal{T}}\mathbf{u}^i_{\mathcal{T}} - N_{\mathcal{T}}(\mathbf{u}^i_{\mathcal{T}})\right\}\\
&\quad + (1-\theta)\left\{\mathbf{f}_{\mathcal{T}}(t_{i-1}) - \nu A_{\mathcal{T}}\mathbf{u}^{i-1}_{\mathcal{T}} - N_{\mathcal{T}}(\mathbf{u}^{i-1}_{\mathcal{T}})\right\}\\
B^T_{\mathcal{T}}\mathbf{u}^i_{\mathcal{T}} &= 0 \quad ,1 \le i \le N,
\end{aligned}$$

where $I_{\mathcal{T}} : H^1_0(\Omega)^n \to X(\mathcal{T})$ denotes a suitable interpolation operator (cf. §I.2.11 (p. 27)).

The equation for $\mathbf{u}^i_{\mathcal{T}}$, $p^i_{\mathcal{T}}$ can be written in the form

$$\begin{aligned}
\frac{1}{\tau_i}\mathbf{u}^i_{\mathcal{T}} + \theta\nu A_{\mathcal{T}}\mathbf{u}^i_{\mathcal{T}} + B_{\mathcal{T}}p^i_{\mathcal{T}} + \theta N_{\mathcal{T}}(\mathbf{u}^i_{\mathcal{T}}) &= \mathbf{g}^i\\
B^T_{\mathcal{T}}\mathbf{u}^i_{\mathcal{T}} &= 0
\end{aligned}$$

with

$$\mathbf{g}^i = \frac{1}{\tau_i}\mathbf{u}^{i-1}_{\mathcal{T}} + \theta\mathbf{f}_{\mathcal{T}}(t_i) + (1-\theta)\left\{\mathbf{f}_{\mathcal{T}}(t_{i-1}) - \nu A_{\mathcal{T}}\mathbf{u}^{i-1}_{\mathcal{T}} - N_{\mathcal{T}}(\mathbf{u}^{i-1}_{\mathcal{T}})\right\}.$$

Thus it is a discrete version of a stationary incompressible Navier-Stokes equation. Hence the methods of §III.2 (p. 85) can be used for its solution.

Due to the condition number of $O(h^{-2})$ of the stiffness matrix $A_{\mathcal{T}}$, one *must* choose the parameter $\theta \ge \frac{1}{2}$. Due to the singularity of the solution of the Navier-Stokes equations at time 0, it is recommended to first perform a few implicit Euler steps, i.e. $\theta = 1$, and then to switch to the Crank-Nicolson scheme, i.e. $\theta = \frac{1}{2}$.

The main drawback of the method of lines is the fixed (w.r.t. time) spatial mesh. Thus singularities, which move through the domain Ω in the course of time, cannot be resolved adaptively. This either leads

to a very poor approximation or to an enormous overhead due to an excessively fine spatial mesh.

IV.1.5. Rothe's method. In the method of lines we first discretize in space and then in time. This order is reversed in *Rothe's method*. We first apply a θ-scheme or a Runge-Kutta method to the instationary Navier-Stokes equations. This leaves us in every time-step with a stationary Navier-Stokes equation.

For the θ-scheme, e.g., we obtain the problems

$$\begin{aligned} \frac{1}{\tau_i}\mathbf{u}^i - \nu\theta\Delta\mathbf{u}^i & \\ +\theta(\mathbf{u}^i\cdot\nabla)\mathbf{u}^i - \operatorname{grad} p^i = \theta\mathbf{f}(\cdot,t_i) + \frac{1}{\tau_i}\mathbf{u}^{i-1} & \\ + (1-\theta)\{\mathbf{f}(\cdot,t_{i-1}) - \nu\Delta\mathbf{u}^{i-1} & \\ + (\mathbf{u}^{i-1}\cdot\nabla)\mathbf{u}^{i-1}\} & \quad \text{in } \Omega \\ \operatorname{div}\mathbf{u}^i = 0 & \quad \text{in } \Omega \\ \mathbf{u}^i = 0 & \quad \text{on } \Gamma. \end{aligned}$$

The stationary Navier-Stokes equations are then discretized in space.

The main difference to the method of lines is the possibility to choose a different partition and spatial discretization at each time-level. This obviously has the advantage that we may apply an adaptive mesh-refinement on each time-level separately. The main drawback of Rothe's method is the lack of a mathematically sound step-size control for the temporal discretization. If we always use the same spatial mesh and the same spatial discretization, Rothe's method yields the same discrete solution as the method of lines.

IV.1.6. Space-time finite elements. This approach circumvents the drawbacks of the method of lines and of Rothe's method. When combined with a suitable space-time adaptivity as described in the next section, it can be viewed as a Rothe's method with a mathematically sound step-size control in space and time.

To describe the method we choose as in §IV.1.3 (p. 96) an integer $N \geq 1$ and intermediate times $0 = t_0 < t_1 < \ldots < t_N = T$ and set $\tau_i = t_i - t_{i-1}$, $1 \leq i \leq N$. With each time t_i we associate a partition $\mathcal{T}_i$ of the spatial domain Ω and corresponding finite element spaces $X(\mathcal{T}_i)$ and $Y(\mathcal{T}_i)$ for the velocity and pressure, respectively. *These spaces must satisfy the inf-sup condition of §II.2.6 (p. 36).* We denote by $\mathbf{u}^i_{\mathcal{T}} \in X(\mathcal{T}_i)$ and $p^i_{\mathcal{T}} \in Y(\mathcal{T}_i)$ the approximations of the velocity and pressure, respectively at time t_i. The discrete problem is then given by

Find $\mathbf{u}_{\mathcal{T}}^0 \in X(\mathcal{T}_0)$ such that

$$\int_\Omega \mathbf{u}_{\mathcal{T}}^0 \cdot \mathbf{v}_{\mathcal{T}}^0 \mathrm{d}x = \int_\Omega \mathbf{u}_0 \cdot \mathbf{v}_{\mathcal{T}}^0 \mathrm{d}x$$

for all $\mathbf{v}_{\mathcal{T}}^0 \in X(\mathcal{T}_0)$ and determine $\mathbf{u}_{\mathcal{T}}^i \in X(\mathcal{T}_i)$ and $p_{\mathcal{T}}^i \in Y(\mathcal{T}_i)$ for $i = 1, \ldots, N$ successively such that

$$\begin{aligned}
&\frac{1}{\tau_i} \int_\Omega \mathbf{u}_{\mathcal{T}}^i \cdot \mathbf{v}_{\mathcal{T}}^i \mathrm{d}x \\
&+\theta\nu \int_\Omega \nabla \mathbf{u}_{\mathcal{T}}^i : \nabla \mathbf{v}_{\mathcal{T}}^i \mathrm{d}x \\
&- \int_\Omega p_{\mathcal{T}}^i \operatorname{div} \mathbf{v}_{\mathcal{T}}^i \mathrm{d}x \\
+\Theta \int_\Omega [(\mathbf{u}_{\mathcal{T}}^i \cdot \nabla)\mathbf{u}_{\mathcal{T}}^i] \cdot \mathbf{v}_{\mathcal{T}}^i \mathrm{d}x &= \frac{1}{\tau_i} \int_\Omega \mathbf{u}_{\mathcal{T}}^{i-1} \cdot \mathbf{v}_{\mathcal{T}}^i \mathrm{d}x \\
&\quad + \theta \int_\Omega \mathbf{f}(x, t_i) \cdot \mathbf{v}_{\mathcal{T}}^i \mathrm{d}x \\
&\quad + (1-\theta) \int_\Omega \mathbf{f}(x, t_{i-1}) \cdot \mathbf{v}_{\mathcal{T}}^i \mathrm{d}x \\
&\quad + (1-\theta)\nu \int_\Omega \nabla \mathbf{u}_{\mathcal{T}}^{i-1} : \nabla \mathbf{v}_{\mathcal{T}}^i \mathrm{d}x \\
&\quad + (1-\Theta) \int_\Omega [(\mathbf{u}_{\mathcal{T}}^{i-1} \cdot \nabla)\mathbf{u}_{\mathcal{T}}^{i-1}] \cdot \mathbf{v}_{\mathcal{T}}^i \mathrm{d}x \\
\int_\Omega q_{\mathcal{T}}^i \operatorname{div} \mathbf{u}_{\mathcal{T}}^i \mathrm{d}x &= 0
\end{aligned}$$

holds for all $\mathbf{v}_{\mathcal{T}}^i \in X(\mathcal{T}_i)$ and $q_{\mathcal{T}}^i \in Y(\mathcal{T}_i)$.

The parameter θ is chosen either equal to $\frac{1}{2}$ or equal to 1. The parameter Θ is chosen either equal to θ or equal to 1.

Note, that $\mathbf{u}_{\mathcal{T}}^0$ is the L^2-projection of $\mathbf{u}_0$ onto $X(\mathcal{T}_0)$ and that the right-hand side of the equation for $\mathbf{u}_{\mathcal{T}}^i$ involves the L^2-projection of functions in $X(\mathcal{T}_{i-1})$ onto $X(\mathcal{T}_i)$. Roughly speaking, these projections make the difference between the space-time finite elements and Rothe's method. Up to these projections one also recovers the method of lines when all partitions and finite element spaces are fixed with respect to time.

On each time-level we have to solve a discrete version of a stationary Navier-Stokes equation. This is done with the methods of §III.2 (p. 85).

IV.1.7. The transport-diffusion algorithm. This method is based on the observation that due to the transport theorem of §I.1.3 (p. 8) the term

$$\frac{\partial \mathbf{u}}{\partial t} + (\mathbf{u} \cdot \nabla)\mathbf{u}$$

is the material derivative along the trajectories of the flow. Therefore

$$\Big(\frac{\partial \mathbf{u}}{\partial t} + (\mathbf{u}\cdot\nabla)\mathbf{u}\Big)(x,t)$$

is approximated by the backward difference

$$\frac{\mathbf{u}(x,t) - \mathbf{u}(y(t-\tau),t-\tau)}{\tau}$$

where $s \mapsto y(s)$ is the trajectory which passes at time t through the point x. The remaining terms are approximated in a standard finite element way.

To describe the algorithm more precisely, we retain the notations of the previous sub-section and introduce in addition on each partition $\mathcal{T}_i$ a quadrature formula

$$\int_\Omega \varphi \mathrm{d}x \approx \sum_{x\in\mathcal{Q}_i} \mu_x \varphi(x)$$

which is exact at least for piecewise constant functions. The most important examples are given by

$$\begin{aligned}
&\mathcal{Q}_i \text{ the barycentres } x_K \text{ of the elements in } \mathcal{T}_i, && \mu_{x_K} = |K|,\\
&\mathcal{Q}_i \text{ the vertices } x \text{ in } \mathcal{T}_i, && \mu_x = \frac{|\omega_x|}{3},\\
&\mathcal{Q}_i \text{ the midpoints } x_E \text{ of the edges in } \mathcal{T}_i, && \mu_{x_E} = \frac{|\omega_E|}{3}.
\end{aligned}$$

Here, $|\omega|$ denotes the area, if $n=2$, or the volume, if $n=3$, of a set $\omega \subset \mathbb{R}^n$ and ω_x and ω_E are the unions of all elements that share the vertex x or the edge E respectively (cf. Figures I.2.2 (p. 24) and I.2.6 (p. 28)).

With these notations the *transport-diffusion algorithm* is given by:

Find $\mathbf{u}^0_{\mathcal{T}} \in X(\mathcal{T}_0)$ such that

$$\sum_{x\in\mathcal{Q}_0} \mu_x \mathbf{u}^0_{\mathcal{T}}(x)\cdot\mathbf{v}^0_{\mathcal{T}}(x) = \sum_{x\in\mathcal{Q}_0} \mu_x \mathbf{u}_0(x)\cdot\mathbf{v}^0_{\mathcal{T}}(x)$$

for all $\mathbf{v}^0_{\mathcal{T}} \in X(\mathcal{T}_0)$. For $i=1,\dots,N$ successively, solve the initial value problems (*transport step*)

$$\begin{aligned}
&\frac{d\mathbf{y}_x(t)}{dt} = \mathbf{u}^{i-1}_{\mathcal{T}}(\mathbf{y}_x(t),t) && \text{for } t_{i-1} < t < t_i\\
&\mathbf{y}_x(t_i) = x
\end{aligned}$$

for all $x \in \mathcal{Q}_i$ and find $\mathbf{u}_{\mathcal{T}}^i \in X(\mathcal{T}_i)$ and $p_{\mathcal{T}}^i \in Y(\mathcal{T}_i)$ such that (*diffusion step*)

$$\begin{aligned}
&\frac{1}{\tau_i}\sum_{x\in\mathcal{Q}_i}\mu_x \mathbf{u}_{\mathcal{T}}^i(x)\cdot\mathbf{v}_{\mathcal{T}}^i(x)\\
&\quad+\theta\nu\int_\Omega \nabla\mathbf{u}_{\mathcal{T}}^i : \nabla\mathbf{v}_{\mathcal{T}}^i \mathrm{d}x\\
&\qquad-\int_\Omega p_{\mathcal{T}}^i \operatorname{div}\mathbf{v}_{\mathcal{T}}^i \mathrm{d}x = \frac{1}{\tau_i}\sum_{x\in\mathcal{Q}_i}\mu_x \mathbf{u}_{\mathcal{T}}^{i-1}(\mathbf{y}_x(t_{i-1}))\cdot\mathbf{v}_{\mathcal{T}}^i(x)\\
&\qquad\qquad+\theta\int_\Omega \mathbf{f}(x,t_i)\cdot\mathbf{v}_{\mathcal{T}}^i \mathrm{d}x\\
&\qquad\qquad+(1-\theta)\int_\Omega \mathbf{f}(x,t_{i-1})\cdot\mathbf{v}_{\mathcal{T}}^i \mathrm{d}x\\
&\qquad\qquad+(1-\theta)\nu\int_\Omega \nabla\mathbf{u}_{\mathcal{T}}^{i-1} : \nabla\mathbf{v}_{\mathcal{T}}^i \mathrm{d}x\\
&\qquad\int_\Omega q_{\mathcal{T}}^i \operatorname{div}\mathbf{u}_{\mathcal{T}}^i \mathrm{d}x = 0
\end{aligned}$$

holds for all $\mathbf{v}_{\mathcal{T}}^i \in X(\mathcal{T}_i)$ and $q_{\mathcal{T}}^i \in Y(\mathcal{T}_i)$.

In contrast to the previous algorithms we now only have to solve discrete Stokes problems at the different time-levels. This is the reward for the necessity to solve the initial value problems for the $\mathbf{y}_x$ and to evaluate functions in $X(\mathcal{T}_i)$ at the points $\mathbf{y}_x(t_{i-1})$ which are not nodal degrees of freedom.

IV.2. Space-time adaptivity

IV.2.1. Overview. When compared to adaptive discretizations of stationary problems an adaptive discretization scheme for instationary problems requires some additional features:

- We need an error estimator which gives upper and lower bounds on the total error, i.e. on both the errors introduced by the spatial and the temporal discretization.
- The error estimator must allow a distinction between the temporal and spatial part of the error.
- We need a strategy for adapting the time-step size.
- The refinement strategy for the spatial discretization must also allow the coarsening of elements in order to take account of singularities which move through the spatial domain in the course of time.

For the Navier-Stokes equations some special difficulties arise in addition:

- The velocities are not exactly solenoidal. This introduces an additional consistency error which must be properly estimated and balanced.
- The convection term may be dominant. Therefore the a posteriori error estimators must be *robust*, i.e. the ratio of upper and lower error bounds must stay bounded uniformly with respect to large values of $Re\|\mathbf{u}\|_1$ or $\frac{1}{\nu}\|\mathbf{u}\|_1$.

IV.2.2. A residual a posteriori error estimator. The error estimator now consists of two components: a spatial part and a temporal one. The spatial part is a straightforward generalization of the residual estimator of §III.2.2 (p. 86) for the stationary Navier-Stokes equation. With the notations of §IV.1.6 (p. 100) the spatial contribution on time-level i is given by

$$\begin{aligned}\eta_h^i = \Bigg\{ \sum_{K\in\mathcal{T}_i} h_K^2 \|\theta\mathbf{f}(\cdot,t_i) + (1-\theta)\mathbf{f}(\cdot,t_{i-1}) - \frac{\mathbf{u}_\mathcal{T}^i - \mathbf{u}_\mathcal{T}^{i-1}}{\tau_i} \\ + \theta\nu\Delta\mathbf{u}_\mathcal{T}^i + (1-\theta)\Delta\nu\mathbf{u}_\mathcal{T}^{i-1} - \nabla p_\mathcal{T}^i \\ - \Theta(\mathbf{u}_\mathcal{T}^i\cdot\nabla)\mathbf{u}_\mathcal{T}^i - (1-\Theta)(\mathbf{u}_\mathcal{T}^{i-1}\cdot\nabla)\mathbf{u}_\mathcal{T}^{i-1}\|_{L^2(K)}^2 \\ + \sum_{K\in\mathcal{T}_i} \|\operatorname{div}\mathbf{u}_\mathcal{T}^i\|_{L^2(K)}^2 \\ + \sum_{E\in\mathcal{E}_i} h_E \|\mathbb{J}_E(\mathbf{n}_E\cdot(\theta\nu\nabla\mathbf{u}_\mathcal{T}^i + (1-\theta)\nu\nabla\mathbf{u}_\mathcal{T}^{i-1} \\ - p_\mathcal{T}^i\underline{\mathbf{I}}))\|_{L^2(E)}^2 \Bigg\}^{\frac{1}{2}}.\end{aligned}$$

Recall that $\mathcal{E}_i$ denotes the set of all interior edges, if $n=2$, respectively interior faces, if $n=3$, in $\mathcal{T}_i$.

In order to obtain a robust estimator, we have to invest some work in the computation of the temporal part of the estimator. For this part we have to solve on each time-level the following discrete Poisson equations:

Find $\widetilde{\mathbf{u}}_\mathcal{T}^i \in S_0^{1,0}(\mathcal{T}_i)^n$ such that

$$\nu\int_\Omega \nabla\widetilde{\mathbf{u}}_\mathcal{T}^i : \nabla\mathbf{v}_\mathcal{T}^i \mathrm{d}x$$

$$= \int_\Omega [((\Theta \mathbf{u}_\mathcal{T}^i + (1-\Theta)\mathbf{u}_\mathcal{T}^{i-1}) \cdot \nabla)(\mathbf{u}_\mathcal{T}^i - \mathbf{u}_\mathcal{T}^{i-1})] \cdot \mathbf{v}_\mathcal{T}^i \mathrm{d}x$$

holds for all $\mathbf{v}_\mathcal{T}^i \in S_0^{1,0}(\mathcal{T}_i)^n$.

The temporal contribution of the error estimator on time-level i then is given by

$$\eta_\tau^i = \Bigg\{ \nu |\mathbf{u}_\mathcal{T}^i - \mathbf{u}_\mathcal{T}^{i-1}|_1^2 + \nu |\widetilde{\mathbf{u}}_\mathcal{T}^i|_1^2 + \sum_{K \in \mathcal{T}_i} h_K^2 \| ((\Theta \mathbf{u}_\mathcal{T}^i + (1-\Theta)\mathbf{u}_\mathcal{T}^{i-1}) \cdot \nabla)(\mathbf{u}_\mathcal{T}^i - \mathbf{u}_\mathcal{T}^{i-1}) + \nu \Delta \widetilde{\mathbf{u}}_\mathcal{T}^i \|_{L^2(K)}^2 + \sum_{E \in \mathcal{E}_i} h_E \| \mathbb{J}_E(\nu \cdot \nabla \widetilde{\mathbf{u}}_\mathcal{T}^i) \|_{L^2(E)}^2 \Bigg\}^{1/2}.$$

With these ingredients the residual a posteriori error estimator for the instationary incompressible Navier-Stokes equations takes the form

$$\left\{ \sum_{i=1}^N \tau_i \left[\left(\eta_h^i\right)^2 + \left(\eta_\tau^i\right)^2 \right] \right\}^{1/2}.$$

IV.2.3. Time adaptivity. Assume that we have solved the discrete problem up to time-level $i-1$ and that we have computed the error estimators η_h^{i-1} and η_τ^{i-1}. Then we set

$$t_i = \begin{cases} \min\{T, t_{i-1} + \tau_{i-1}\} & \text{if } \eta_\tau^{i-1} \approx \eta_h^{i-1}, \\ \min\{T, t_{i-1} + 2\tau_{i-1}\} & \text{if } \eta_\tau^{i-1} \le \frac{1}{2}\eta_h^{i-1}. \end{cases}$$

In the first case we retain the previous time-step; in the second case we try a larger time step.

Next, we solve the discrete problem on time-level i with the current value of t_i and compute the error estimators η_h^i and η_τ^i.

If $\eta_\tau^i \approx \eta_h^i$, we accept the current time-step and continue with the space adaptivity, which is described in the next sub-section.

If $\eta_\tau^i \ge 2\eta_h^i$, we reject the current time-step. We replace t_i by $\frac{1}{2}(t_{i-1} + t_i)$ and repeat the solution of the discrete problem on time-level i and the computation of the error estimators.

The described strategy obviously aims at balancing the two contributions η_h^i and η_τ^i of the error estimator.

IV.2.4. Space adaptivity. For time-dependent problems the spatial adaptivity must also allow for a local mesh coarsening (cf. [11, §III.1.5]). Hence, the marking strategies of §II.7.5 (p. 69) must be modified accordingly.

Assume that we have solved the discrete problem on time-level i with an actual time-step τ_i and an actual partition $\mathcal{T}_i$ of the spatial domain Ω and that we have computed the estimators η_h^i and η_τ^i. Moreover, suppose that we have accepted the current time-step and want to optimize the partition $\mathcal{T}_i$.

We may assume that $\mathcal{T}_i$ currently is the finest partition in a hierarchy $\mathcal{T}_i^0, \dots, \mathcal{T}_i^\ell$ of nested, successively refined partitions, i.e. $\mathcal{T}_i = \mathcal{T}_i^\ell$ and $\mathcal{T}_i^j$ is a (local) refinement of $\mathcal{T}_i^{j-1}$, $1 \le j \le \ell$.

Now, we go back m generations in the grid-hierarchy to the partition $\mathcal{T}_i^{\ell-m}$. Due to the nestedness of the partitions, each element $K \in \mathcal{T}_i^{\ell-m}$ is the union of several elements $K' \in \mathcal{T}_i$. Each K' gives a contribution to η_h^i. We add these contributions and thus obtain for every $K \in \mathcal{T}_i^{\ell-m}$ an error estimator η_K. With these estimators we then perform M steps of the marking strategy of §II.7.5 (p. 69). This yields a new partition $\mathcal{T}_i^{\ell-m+M}$ which usually is different from $\mathcal{T}_i$. We replace $\mathcal{T}_i$ by this partition, solve the corresponding discrete problem on time-level i and compute the new error estimators η_h^i and η_τ^i.

If the newly calculated error estimators satisfy $\eta_h^i \approx \eta_\tau^i$, we accept the current partition $\mathcal{T}_i$ and proceed with the next time-level.

If $\eta_h^i \ge 2\eta_\tau^i$, we successively refine the partition $\mathcal{T}_i$ as described in §II.7 (p. 62) with the η_h^i as error estimators until we arrive at a partition which satisfies $\eta_h^i \approx \eta_\tau^i$. When this goal is achieved we accept the spatial discretization and proceed with the next time-level.

Typical values for the parameters m and M are $1 \le m \le 3$ and $m \le M \le m+2$.

IV.3. Discretization of compressible and inviscid problems

IV.3.1. Systems in divergence form. In this section we consider problems of the following form:
Given a domain $\Omega \subset \mathbb{R}^n$ with $n = 2$ or $n = 3$, an integer $m \ge n$, a vector field $\mathbf{g} : \mathbb{R}^m \times \Omega \times (0,\infty) \to \mathbb{R}^m$, a vector field $\mathbf{M} : \mathbb{R}^m \to \mathbb{R}^m$, a tensor field $\underline{\mathbf{F}} : \mathbb{R}^m \to \mathbb{R}^{m\times n}$, and a vector field $\mathbf{U}_0 : \Omega \to \mathbb{R}^m$, we are looking for a vector field $\mathbf{U} : \Omega \times (0,\infty) \to \mathbb{R}^m$ such that

$$\frac{\partial \mathbf{M}(\mathbf{U})}{\partial t} + \operatorname{div} \underline{\mathbf{F}}(\mathbf{U}) = \mathbf{g}(\mathbf{U}, x, t) \qquad \text{in } \Omega \times (0,\infty)$$

$$\mathbf{U}(\cdot,0) = \mathbf{U}_0 \qquad \text{in } \Omega.$$

Such a problem, which has to be complemented with suitable boundary conditions, is called a *system (of differential equations) in divergence form*.

Note, that the divergence is taken row-wise, i.e.

$$\operatorname{div} \underline{\mathbf{F}}(\mathbf{U}) = \left(\sum_{j=1}^{n} \frac{\partial \underline{\mathbf{F}}(\mathbf{U})_{i,j}}{\partial x_j} \right)_{1 \le i \le m}.$$

The tensor field $\underline{\mathbf{F}}$ is called the *flux* of the system. It is often split into an *advective flux* $\underline{\mathbf{F}}_{\text{adv}}$ which contains no derivatives and a *viscous flux* $\underline{\mathbf{F}}_{\text{visc}}$ which contains spatial derivatives, i.e

$$\underline{\mathbf{F}} = \underline{\mathbf{F}}_{\text{adv}} + \underline{\mathbf{F}}_{\text{visc}}.$$

EXAMPLE IV.3.1. The compressible Navier-Stokes equations and the Euler equations (with $\alpha = 0$) of §I.1.9 (p. 12) and §I.1.10 (p. 13) fit into this framework. For both equations we have

$$\begin{aligned} m &= n + 2 \\ \mathbf{U} &= (\rho, \mathbf{v}, e)^T \\ M(\mathbf{U}) &= \begin{pmatrix} \rho \\ \rho\mathbf{v} \\ e \end{pmatrix} \\ \underline{\mathbf{F}}_{\text{adv}}(\mathbf{U}) &= \begin{pmatrix} \rho\mathbf{v} \\ \rho\mathbf{v} \otimes \mathbf{v} + p\underline{\mathbf{I}} \\ e\mathbf{v} + p\mathbf{v} \end{pmatrix} \\ \mathbf{g} &= \begin{pmatrix} 0 \\ \rho\mathbf{f} \\ \mathbf{f} \cdot \mathbf{v} \end{pmatrix}. \end{aligned}$$

For the Navier-Stokes equations the viscous forces yield a viscous flux

$$\underline{\mathbf{F}}_{\text{visc}}(\mathbf{U}) = \begin{pmatrix} 0 \\ \underline{\mathbf{T}} + p\underline{\mathbf{I}} \\ (\underline{\mathbf{T}} + p\underline{\mathbf{I}}) \cdot \mathbf{v} + \sigma \end{pmatrix}.$$

These equations are complemented by the constitutive equations for p, $\underline{\mathbf{T}}$, and σ given in §I.1.8 (p. 11).

IV.3.2. Finite volume schemes. Finite volume schemes are particularly suited for the discretization of systems in divergence form.

To describe the idea in its simplest form, we choose a time-step $\tau > 0$ and a partition $\mathcal{T}$ of the domain Ω. *The partition may consist of arbitrary polyhedrons. From §I.2.7 (p. 21) we only retain the condition that the sub-domains must not overlap.*

In a first step, we fix an i and an element $K \in \mathcal{T}$. Then we integrate the system over the set $K \times [(i-1)\tau, i\tau]$. This yields

$$\int_{(i-1)\tau}^{i\tau} \int_K \frac{\partial \mathbf{M}(\mathbf{U})}{\partial t} \mathrm{d}x\mathrm{d}t + \int_{(i-1)\tau}^{i\tau} \int_K \operatorname{div} \underline{\mathbf{F}}(\mathbf{U}) \mathrm{d}x\mathrm{d}t = \int_{(i-1)\tau}^{i\tau} \int_K \mathbf{g}(\mathbf{U}, x, t) \mathrm{d}x\mathrm{d}t.$$

Using the integration by parts formulae of §I.2.3 (p. 18), we obtain for the left-hand side

$$\int_{(i-1)\tau}^{i\tau} \int_K \frac{\partial \mathbf{M}(\mathbf{U})}{\partial t} \mathrm{d}x\mathrm{d}t = \int_K \mathbf{M}(\mathbf{U}(x, i\tau)) \mathrm{d}x - \int_K \mathbf{M}(\mathbf{U}(x, (i-1)\tau)) \mathrm{d}x$$

$$\int_{(i-1)\tau}^{i\tau} \int_K \operatorname{div} \underline{\mathbf{F}}(\mathbf{U}) \mathrm{d}x\mathrm{d}t = \int_{(i-1)\tau}^{i\tau} \int_{\partial K} \underline{\mathbf{F}}(\mathbf{U}) \cdot \mathbf{n}_K \mathrm{d}S\mathrm{d}t,$$

where $\mathbf{n}_K$ denotes the exterior unit normal of K.

Now, we assume that $\mathbf{U}$ is piecewise constant with respect to space and time. Denoting by $\mathbf{U}_K^i$ and $\mathbf{U}_K^{i-1}$ its constant values on K at times $i\tau$ and $(i-1)\tau$ respectively, we obtain

$$\int_K \mathbf{M}(\mathbf{U}(x, i\tau)) \mathrm{d}x \approx |K| \mathbf{M}(\mathbf{U}_K^i)$$

$$\int_K \mathbf{M}(\mathbf{U}(x, (i-1)\tau)) \mathrm{d}x \approx |K| \mathbf{M}(\mathbf{U}_K^{i-1}),$$

where $|K|$ denotes the area, if $n = 2$, respectively volume, if $n = 3$, of the element K.

Next we approximate the flux term

$$\int_{(i-1)\tau}^{i\tau} \int_{\partial K} \underline{\mathbf{F}}(\mathbf{U}) \cdot \mathbf{n}_K \mathrm{d}S\mathrm{d}t \approx \tau \int_{\partial K} \underline{\mathbf{F}}(\mathbf{U}_K^{i-1}) \cdot \mathbf{n}_K \mathrm{d}S.$$

The right-hand side is approximated by

$$\int_{(i-1)\tau}^{i\tau} \int_K \mathbf{g}(\mathbf{U}, x, t) \mathrm{d}x\mathrm{d}t \approx \tau |K| \mathbf{g}(\mathbf{U}_K^{i-1}, x_K, (i-1)\tau),$$

where x_K denotes a point in K, which is fixed a priori, e.g. its barycentre.

Finally, we approximate the boundary integral for the flux by a *numerical flux*

$$\tau \int_{\partial K} \underline{\mathbf{F}}(\mathbf{U}_K^{i-1}) \cdot \mathbf{n}_K \mathrm{d}S \approx \tau \sum_{\substack{K' \in \mathcal{T} \\ \partial K \cap \partial K' \in \mathcal{E}}} |\partial K \cap \partial K'| \mathbf{F}_{\mathcal{T}}(\mathbf{U}_K^{i-1}, \mathbf{U}_{K'}^{i-1}),$$

where $\mathcal{E}$ denotes the set of all edges, if $n = 2$, respectively faces, if $n = 3$, in $\mathcal{T}$ and where $|\partial K \cap \partial K'|$ is the length respectively area of the common edge or face of K and K'.

With these approximations the simplest *finite volume scheme* for a system in divergence form is given by

For every element $K \in \mathcal{T}$ compute

$$\mathbf{U}_K^0 = \frac{1}{|K|} \int_K \mathbf{U}_0(x) \mathrm{d}x.$$

For $i = 1, 2, \ldots$ successively compute for all elements $K \in \mathcal{T}$

$$\begin{aligned} \mathbf{M}(\mathbf{U}_K^i) = \mathbf{M}(\mathbf{U}_K^{i-1}) & \\ - \tau \sum_{\substack{K' \in \mathcal{T} \\ \partial K \cap \partial K' \in \mathcal{E}}} & \frac{|\partial K \cap \partial K'|}{|K|} \mathbf{F}_{\mathcal{T}}(\mathbf{U}_K^{i-1}, \mathbf{U}_{K'}^{i-1}) \\ + \tau \mathbf{g}(\mathbf{U}_K^{i-1}, & x_K, (i-1)\tau). \end{aligned}$$

For a concrete finite volume scheme we of course have to specify

- the partition $\mathcal{T}$ and
- the numerical flux $\mathbf{F}_{\mathcal{T}}$.

This will be the subject of the following sections.

Remark IV.3.2. In practice one works with a variable time-step and variable partitions. To this end one chooses an increasing sequence $0 = t_0 < t_1 < t_2 < \ldots$ of times and associates with each time t_i a partition $\mathcal{T}_i$ of Ω. Then τ is replaced by $\tau_i = t_i - t_{i-1}$ and K and K' are elements in $\mathcal{T}_{i-1}$ or $\mathcal{T}_i$. Moreover one has to furnish an interpolation operator which maps piecewise constant functions with respect to one partition to piecewise constant functions with respect to another partition.

IV.3.3. Construction of the partitions. An obvious possibility is to construct a finite volume partition $\mathcal{T}$ in the same way as a finite element partition. In practice, however, one prefers so-called *dual meshes*.

To describe the idea, we consider the two-dimensional case, i.e. $n = 2$. We start from a standard finite element partition $\widetilde{\mathcal{T}}$ which satisfies the conditions of §I.2.7 (p. 21). Then we subdivide each element $\widetilde{K} \in \widetilde{\mathcal{T}}$ into smaller elements by either

- drawing the perpendicular bisectors at the midpoints of edges of $\widetilde{K}$ (cf. Figure IV.3.1) or by
- connecting the barycentre of $\widetilde{K}$ with its midpoints of edges (cf. Figure IV.3.2).

Then the elements in $\mathcal{T}$ consist of the unions of all small elements that share a common vertex in the partition $\widetilde{\mathcal{T}}$.

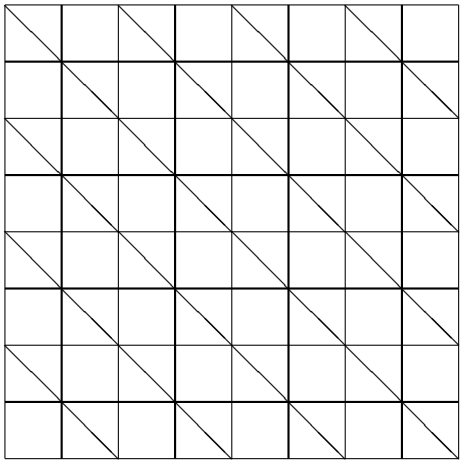

FIGURE IV.3.1. Dual mesh (thick lines) via perpendicular bisectors of primal mesh (thin lines)

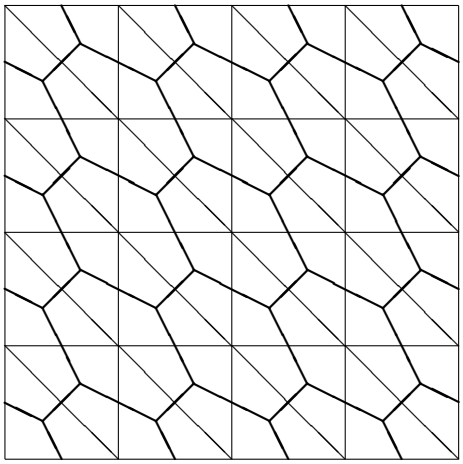

FIGURE IV.3.2. Dual mesh (thick lines) via barycentres of primal mesh (thin lines)

Thus the elements in $\mathcal{T}$ can be associated with the vertices in $\widetilde{\mathcal{N}}$. Moreover, we may associate with each edge in $\mathcal{E}$ exactly two vertices in $\widetilde{\mathcal{N}}$ such that the line connecting these vertices intersects the given edge (cf. Figure IV.3.3).

The first construction has the advantage that this intersection is orthogonal. Yet this construction also has some disadvantages which are not present with the second construction:

- The perpendicular bisectors of a triangle may intersect in a point outside the triangle. The intersection point is within the triangle only if its largest angle is at most a right one.
- The perpendicular bisectors of a quadrilateral may not intersect at all. They intersect in a common point inside the quadrilateral only if it is a rectangle.
- The first construction has no three dimensional analogue.

IV.3.4. Construction of the numerical fluxes. In order to construct the numerical flux $\mathbf{F}_{\mathcal{T}}(\mathbf{U}_K^{i-1}, \mathbf{U}_{K'}^{i-1})$ corresponding to the common boundary of two elements K and K' in a finite volume partition, we split $\partial K \cap \partial K'$ into *straight* edges, if $n = 2$, or *plane* faces, if $n = 3$.

Consider such an edge respectively face E. To simplify the notation, we denote the adjacent elements by K_1 and K_2 and write $\mathbf{U}_1$ and $\mathbf{U}_2$ instead of $\mathbf{U}_{K_1}^{i-1}$ and $\mathbf{U}_{K_2}^{i-1}$, respectively.

We assume that $\mathcal{T}$ is the dual mesh to a finite element partition $\widetilde{\mathcal{T}}$ as described in §IV.3.3 (p. 109). Denote by x_1 and x_2 the two vertices in $\widetilde{\mathcal{N}}$ such that their connecting line intersects E (cf. Figure IV.3.3).

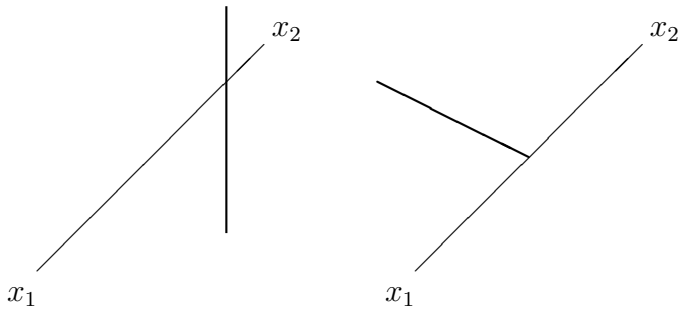

FIGURE IV.3.3. Examples of a common edge E (thick lines) of two volumes and corresponding nodes x_1 and x_2 of the underlying primal mesh together with their connecting line (thin lines)

In order to construct the contribution of E to the numerical flux $\mathbf{F}_{\mathcal{T}}(\mathbf{U}_1, \mathbf{U}_2)$, we split the flux in an advective part $\mathbf{F}_{\mathcal{T}_{\text{adv}}}(\mathbf{U}_1, \mathbf{U}_2)$ and a viscous part $\mathbf{F}_{\mathcal{T}_{\text{visc}}}(\mathbf{U}_1, \mathbf{U}_2)$ similar to the splitting of the analytical flux $\underline{\mathbf{F}}$.

For the viscous part of the numerical flux we introduce a local co-ordinate system $\eta_1, \ldots, \eta_n$ such that the direction η_1 is parallel to the direction $\overline{x_1 x_2}$ and such that the other directions are tangential to E. Note, that in general $\overline{x_1 x_2}$ will not be orthogonal to E. Then we express all derivatives in $\underline{\mathbf{F}}_{\text{visc}}$ in terms of the new co-ordinate system,

suppress all derivatives not involving η_1, and approximate derivatives with respect to η_1 by difference quotients of the form $\frac{\varphi_1-\varphi_2}{\|x_1-x_2\|_2}$. Here, $\|x_1-x_2\|_2$ denotes the Euclidean distance of the two points x_1 and x_2.

For the advective part of the numerical flux we need some additional notations. Consider an arbitrary vector $\mathbf{V} \in \mathbb{R}^m$. Denote by $D(\underline{\mathbf{F}}_{\text{adv}}(\mathbf{V}) \cdot \mathbf{n}_{K_1}) \in \mathbb{R}^{m\times m}$ the derivative of $\underline{\mathbf{F}}_{\text{adv}}(\mathbf{V}) \cdot \mathbf{n}_{K_1}$ with respect to $\mathbf{V}$. One can prove that for many systems in divergence form, including the compressible Navier-Stokes and Euler equations, this matrix can be diagonalized, i.e. there is an invertible $m \times m$ matrix $Q(\mathbf{V})$ and a diagonal $m \times m$ matrix $\Delta(\mathbf{V})$ such that

$$Q(\mathbf{V})^{-1} D(\underline{\mathbf{F}}_{\text{adv}}(\mathbf{V}) \cdot \mathbf{n}_{K_1}) Q(\mathbf{V}) = \Delta(\mathbf{V}).$$

The entries of $\Delta(\mathbf{V})$ are the eigenvalues of $D(\underline{\mathbf{F}}_{\text{adv}}(\mathbf{V}) \cdot \mathbf{n}_{K_1})$.

For any real number z we set

$$z^+ = \max\{z, 0\}\ ,\ z^- = \min\{z, 0\}$$

and define

$$\Delta(\mathbf{V})^{\pm} = \begin{pmatrix} \Delta(\mathbf{V})^{\pm}_{11} & 0 & \dots & 0 \\ 0 & \Delta(\mathbf{V})^{\pm}_{22} & \dots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \dots & \Delta(\mathbf{V})^{\pm}_{mm} \end{pmatrix}$$

and

$$C(\mathbf{V})^{\pm} = Q(\mathbf{V}) \Delta(\mathbf{V})^{\pm} Q(\mathbf{V})^{-1}.$$

With these notations, the *Steeger-Warming approximation* of the advective flux is given by

$$\boxed{\mathbf{F}_{\mathcal{T}}(\mathbf{U}_1, \mathbf{U}_2) = C(\mathbf{U}_1)^+ \mathbf{U}_1 + C(\mathbf{U}_2)^- \mathbf{U}_2.}$$

Another popular numerical flux is the *van Leer approximation* which is given by

$$\boxed{\begin{aligned} &\mathbf{F}_{\mathcal{T}}(\mathbf{U}_1, \mathbf{U}_2) \\ &= \Big[\frac{1}{2} C(\mathbf{U}_1) + C(\frac{1}{2}(\mathbf{U}_1 + \mathbf{U}_2))^+ - C(\frac{1}{2}(\mathbf{U}_1 + \mathbf{U}_2))^-\Big] \mathbf{U}_1 \\ &+ \Big[\frac{1}{2} C(\mathbf{U}_2) - C(\frac{1}{2}(\mathbf{U}_1 + \mathbf{U}_2))^+ + C(\frac{1}{2}(\mathbf{U}_1 + \mathbf{U}_2))^-\Big] \mathbf{U}_2. \end{aligned}}$$

In both schemes one has to compute the derivatives $D\underline{\mathbf{F}}_{\text{adv}}(\mathbf{V}) \cdot \mathbf{n}_{K_1}$ and their eigenvalues and their eigenvectors for suitable values of $\mathbf{V}$. At first sight the van Leer scheme seems to be more costly than the Steeger-Warming scheme since it requires three evaluations of $C(\mathbf{V})$ instead of two. For the compressible Navier-Stokes and Euler equations, however,

this can be reduced to one evaluation since for these equations $\underline{\mathbf{F}}_{\mathrm{adv}}(\mathbf{V}) \cdot \mathbf{n}_{K_1} = C(\mathbf{V})\mathbf{V}$ holds for all $\mathbf{V}$.

EXAMPLE IV.3.3. Consider the *Burger's equation*

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} = 0$$

which, in a certain sense, is a one dimensional counterpart of the Euler equations and which is a system in divergence form with

$$m = n = 1, \qquad \mathbf{u} = u, \quad \mathbf{M}(\mathbf{U}) = u,$$
$$\underline{\mathbf{F}}_{\mathrm{adv}}(u) = \frac{1}{2}u^2, \qquad \underline{\mathbf{F}}_{\mathrm{visc}}(\mathbf{U}) = 0, \qquad \mathbf{g}(\mathbf{U}) = 0.$$

For this equation the Steger-Warming scheme takes the form

$$\underline{\mathbf{F}}_{\mathcal{T},\mathrm{adv}}(u_1, u_2) = \begin{cases} u_1^2 & \text{if } u_1 \geq 0, u_2 \geq 0 \\ u_1^2 + u_2^2 & \text{if } u_1 \geq 0, u_2 \leq 0 \\ u_2^2 & \text{if } u_1 \leq 0, u_2 \leq 0 \\ 0 & \text{if } u_1 \leq 0, u_2 \geq 0 \end{cases}$$

while the van Leer scheme reads

$$\underline{\mathbf{F}}_{\mathcal{T},\mathrm{adv}}(u_1, u_2) = \begin{cases} u_1^2 & \text{if } u_1 \geq -u_2 \\ u_2^2 & \text{if } u_1 \leq -u_2. \end{cases}$$

IV.3.5. Relation to finite element methods. The fact that the elements of a dual mesh can be associated with the vertices of a finite element partition gives a link between finite volume and finite element methods:
Consider a function φ that is piecewise constant on the dual mesh $\mathcal{T}$, i.e. $\varphi \in S^{0,-1}(\mathcal{T})$. With φ we associate a continuous piecewise linear function $\Phi \in S^{1,0}(\widetilde{\mathcal{T}})$ corresponding to the finite element partition $\widetilde{\mathcal{T}}$ such that $\Phi(x_K) = \varphi_K$ for the vertex $x_K \in \widetilde{\mathcal{N}}$ corresponding to $K \in \mathcal{T}$.

This link sometimes considerably simplifies the analysis of finite volume methods. For example it suggests a very simple and natural approach to a posteriori error estimation and mesh adaptivity for finite volume methods:

- Given the solution φ of the finite volume scheme compute the corresponding finite element function Φ.
- Apply a standard a posteriori error estimator to Φ.
- Given the error estimator apply a standard mesh refinement strategy to the finite element mesh $\widetilde{\mathcal{T}}$ and thus construct a new, locally refined partition $\widehat{\mathcal{T}}$.
- Use $\widehat{\mathcal{T}}$ to construct a new dual mesh $\mathcal{T}'$. This is the refinement of $\mathcal{T}$.

IV.3.6. Discontinuous Galerkin methods. These methods can be interpreted as a mixture of finite element and finite volume methods. The basic idea of discontinuous Galerkin methods can be described as follows:

- Approximate $\mathbf{U}$ by discontinuous functions which are polynomials with respect to space and time on small space-time cylinders of the form $K \times [(n-1)\tau, n\tau]$ with $K \in \mathcal{T}$.
- For every such cylinder multiply the differential equation by a corresponding test-polynomial and integrate the result over the cylinder.
- Use integration by parts for the flux term.
- Accumulate the contributions of all elements in $\mathcal{T}$.
- Compensate for the illegal integration by parts by adding appropriate jump-terms across the element boundaries.
- Stabilize the scheme in a Petrov-Galerkin way by adding suitable element residuals.

In their simplest form these ideas lead to the following discrete problem:

Compute $\mathbf{U}_\mathcal{T}^0$, the L^2-projection of $\mathbf{U}_0$ onto $S^{k,-1}(\mathcal{T})$.
For $n \geq 1$ successively find $\mathbf{U}_\mathcal{T}^n \in S^{k,-1}(\mathcal{T})$ such that

$$\begin{aligned}
&\sum_{K\in\mathcal{T}} \frac{1}{\tau} \int_K M(\mathbf{U}_\mathcal{T}^n) \cdot \mathbf{V}_\mathcal{T} \mathrm{d}x - \sum_{K\in\mathcal{T}} \int_K \underline{\mathbf{F}}(\mathbf{U}_\mathcal{T}^n) : \nabla \mathbf{V}_\mathcal{T} \mathrm{d}x \\
&\quad + \sum_{E\in\mathcal{E}} \delta_E h_E \int_E \mathbb{J}_E(\mathbf{n}_E \cdot \underline{\mathbf{F}}(\mathbf{U}_\mathcal{T}^n)\mathbf{V}_\mathcal{T}) \mathrm{d}S \\
&\quad + \sum_{K\in\mathcal{T}} \delta_K h_K^2 \int_K \operatorname{div} \underline{\mathbf{F}}(\mathbf{U}_\mathcal{T}^n) \cdot \operatorname{div} \underline{\mathbf{F}}(\mathbf{V}_\mathcal{T}) \mathrm{d}x \\
&= \sum_{K\in\mathcal{T}} \frac{1}{\tau} \int_K M(\mathbf{U}_\mathcal{T}^{n-1}) \cdot \mathbf{V}_\mathcal{T} \mathrm{d}x + \sum_{K\in\mathcal{T}} \int_K \mathbf{g}(\cdot, n\tau) \cdot \mathbf{V}_\mathcal{T} \mathrm{d}x \\
&\quad + \sum_{K\in\mathcal{T}} \delta_K h_K^2 \int_K \mathbf{g}(\cdot, n\tau) \cdot \operatorname{div} \underline{\mathbf{F}}(\mathbf{V}_\mathcal{T}) \mathrm{d}x
\end{aligned}$$

holds for all $\mathbf{V}_\mathcal{T}$.

This discretization can easily be generalized as follows:

- The jump and stabilization terms can be chosen more judiciously.
- The time-step may not be constant.
- The spacial mesh may depend on time.
- The functions $\mathbf{U}_\mathcal{T}$ and $\mathbf{V}_\mathcal{T}$ may be piecewise polynomials of higher order with respect to to time. Then the term

$$\sum_{K\in\mathcal{T}} \int_{(n-1)\tau}^{n\tau} \int_K \frac{\partial M(\mathbf{U}_\mathcal{T})}{\partial t} \cdot \mathbf{V}_\mathcal{T} \mathrm{d}x \mathrm{d}t$$

must be added on the left-hand side and terms of the form

$$\frac{\partial M(\mathbf{U}_{\mathcal{T}})}{\partial t} \cdot \mathbf{V}_{\mathcal{T}}$$

must be added to the element residuals.

Bibliography

[1] M. Ainsworth, J. T. Oden: A Posteriori Error Estimation in Finite Element Analysis. Wiley, 2000

[2] V. Girault, P. A. Raviart: Finite Element Approximation of the Navier-Stokes Equations. Computational Methods in Physics, Springer, Berlin, 2nd edition, 1986

[3] R. Glowinski: Finite Element Methods for Incompressible Viscous Flows. Handbook of Numerical Analysis Vol. IX, Elsevier 2003

[4] E. Godlewski, P. A. Raviart: Numerical Approximation of Hyperbolic Systems of Conservation Laws. Springer, 1996

[5] D. Gunzburger, R. A. Nicolaides: Incompressible CFD - Trends and Advances. Cambridge University Press, 1993

[6] D. Kröner: Numerical Schemes for Conservation Laws. Teubner-Wiley, 1997

[7] R. LeVeque: Finite Volume Methods for Hyperbolic Problems. Springer, 2002

[8] O. Pironneau: Finite Element Methods for Fluids. Wiley, 1989

[9] R. Temam: Navier-Stokes Equations. 3rd edition, North Holland, 1984

[10] R. Verfürth: A Posteriori Error Estimation Techniques for Finite Element Methods, Oxford University Press, Oxford, 2013.

[11] R. Verfürth: Adaptive Finite Element Methods. Lecture Notes, Ruhr-Universität Bochum 2011 (`www.rub.de/num1/files/lectures/AdaptiveFEM.pdf`)

Index