Chapter 5
The Steady-State Navier–Stokes Equations

Remark 5.1. The steady state Navier–Stokes equations. The steady-state or stationary Navier–Stokes equations describe steady-state flows. Such flow fields can be expected in practice if:

- all data of the Navier–Stokes equations (1.23) do not depend on the time,
- the viscosity ν is sufficiently large, or equivalently, the Reynolds number Re is sufficiently small,

see Remark 1.22.

The Navier–Stokes equations are nonlinear. That means, the third difficulty mentioned Remark 1.19 has to be addressed. □

5.1 The Continuous Equations

5.1.1 The Strong Form and the Variational Form

Remark 5.2. Strong form of the steady-state Navier–Stokes equations. The steady-state Navier–Stokes equations are given by

$$\begin{aligned} -\nu\Delta\boldsymbol{u} + (\boldsymbol{u}\cdot\nabla)\boldsymbol{u} + \nabla p &= \boldsymbol{f} \text{ in } \Omega, \\ \nabla\cdot\boldsymbol{u} &= 0 \text{ in } \Omega, \end{aligned} \tag{5.1}$$

where Ω is a bounded domain with Lipschitz boundary. As for the Stokes and Oseen equations, the numerical analysis will be presented for the case of homogeneous Dirichlet boundary conditions $\boldsymbol{u} = \boldsymbol{0}$ on Γ. □

Remark 5.3. Variational form of the steady-state Navier–Stokes equations. For the variational formulation of the steady-state Navier–Stokes equations, the same functions spaces $V = H_0^1(\Omega)$ and $Q = L_0^2(\Omega)$ as for the Stokes and Oseen equations can be used. A variational form of (5.1) is as follows: Find $(\boldsymbol{u}, p) \in V \times Q$ such that

$$\begin{aligned}(\nu\nabla\boldsymbol{u},\nabla\boldsymbol{v}) + ((\boldsymbol{u}\cdot\nabla)\boldsymbol{u},\boldsymbol{v}) - (\nabla\cdot\boldsymbol{v},p) &= \langle\boldsymbol{f},\boldsymbol{v}\rangle_{V',V},\\ -(\nabla\cdot\boldsymbol{u},q) &= 0\end{aligned} \tag{5.2}$$

for all $(\boldsymbol{v},q)\in V\times Q$. The problem can be formulated equivalently as follows: Find $(\boldsymbol{u},p)\in V\times Q$ such that

$$(\nu\nabla\boldsymbol{u},\nabla\boldsymbol{v}) + ((\boldsymbol{u}\cdot\nabla)\boldsymbol{u},\boldsymbol{v}) - (\nabla\cdot\boldsymbol{v},p) + (\nabla\cdot\boldsymbol{u},q) = \langle\boldsymbol{f},\boldsymbol{v}\rangle_{V',V} \tag{5.3}$$

for all $(\boldsymbol{v},q)\in V\times Q$.

By the Sobolov imbeddings $H^1(\Omega)\to L^4(\Omega)$, see (A.16) and (A.17), one has that $\boldsymbol{u},\boldsymbol{v}\in L^4(\Omega)$. Since $\nabla\boldsymbol{u}\in L^2(\Omega)$, it follows from the generalized Hölder inequality that the term $((\boldsymbol{u}\cdot\nabla)\boldsymbol{u},\boldsymbol{v})$ is well defined if $\boldsymbol{u},\boldsymbol{v}\in V$. □

Remark 5.4. The reduced problem in V_{div}. There is also an associated problem in the space V_{div} of weakly divergence-free functions: Find $\boldsymbol{u}\in V_{\text{div}}$ such that

$$(\nu\nabla\boldsymbol{u},\nabla\boldsymbol{v}) + ((\boldsymbol{u}\cdot\nabla)\boldsymbol{u},\boldsymbol{v}) = \langle\boldsymbol{f},\boldsymbol{v}\rangle_{V',V} \quad \forall\,\boldsymbol{v}\in V_{\text{div}}. \tag{5.4}$$

Of course, if $(\boldsymbol{u},p)\in V\times Q$ is a solution of (5.3), $\boldsymbol{u}\in V_{\text{div}}$ and $\boldsymbol{u}$ is a solution of (5.4). The other direction can be proved as for linear saddle point problems, see Section 2.1. The bilinear form which couples velocity and pressure is the same as for the Stokes and the Oseen equations and the spaces are the same, too. Thus, given a solution $\boldsymbol{u}$ of (5.4), there exists a unique pressure $p\in Q$ such that $(\boldsymbol{u},p)$ solves (5.3) if the spaces V and Q satisfy the inf-sup condition (2.14). Note that in the proof of the inf-sup condition, see the proof of Lemma 2.12, the bilinear form which couples the ansatz and test functions of the velocity space does not play any role. □

5.1.2 The Nonlinear Term

Remark 5.5. Different forms of the convective term in (5.1). There are several forms of the convective term of the incompressible Navier–Stokes equations:

$(\boldsymbol{u}\cdot\nabla)\boldsymbol{u}$	:	convective form,
$\nabla\cdot\left(\boldsymbol{u}\boldsymbol{u}^T\right)$	:	divergence form,
$\nabla\cdot\left(\boldsymbol{u}\boldsymbol{u}^T\right)+\frac{1}{2}\nabla\left(\boldsymbol{u}^T\boldsymbol{u}\right)$	:	divergence form with modified pressure,
$(\nabla\times\boldsymbol{u})\times\boldsymbol{u}$	:	rotational form (with modified pressure).

Let $\boldsymbol{u}=(u_1,u_2,u_3)^T$ be weakly differentiable and weakly divergence-free, i.e., $\nabla\cdot\boldsymbol{u}=0$ almost everywhere.

- *Convective form and divergence form.* Then, the convective form and the divergence form are equivalent, since one gets with the product rule

$$
\begin{aligned}
&\nabla \cdot \left(\boldsymbol{u}\boldsymbol{u}^T\right) \\
&= \nabla \cdot \begin{pmatrix} u_1u_1 & u_1u_2 & u_1u_3 \\ u_2u_1 & u_2u_2 & u_2u_3 \\ u_3u_1 & u_3u_2 & u_3u_3 \end{pmatrix} = \begin{pmatrix} \partial_x(u_1u_1) + \partial_y(u_1u_2) + \partial_z(u_1u_3) \\ \partial_x(u_2u_1) + \partial_y(u_2u_2) + \partial_z(u_2u_3) \\ \partial_x(u_3u_1) + \partial_y(u_3u_2) + \partial_z(u_3u_3) \end{pmatrix} \\
&= \begin{pmatrix} (u_1\partial_x + u_2\partial_y + u_3\partial_z)u_1 \\ (u_1\partial_x + u_2\partial_y + u_3\partial_z)u_2 \\ (u_1\partial_x + u_2\partial_y + u_3\partial_z)u_3 \end{pmatrix} + \begin{pmatrix} u_1(\partial_x(u_1) + \partial_y(u_2) + \partial_z(u_3)) \\ u_2(\partial_x(u_1) + \partial_y(u_2) + \partial_z(u_3)) \\ u_3(\partial_x(u_1) + \partial_y(u_2) + \partial_z(u_3)) \end{pmatrix} \\
&= (\boldsymbol{u} \cdot \nabla)\boldsymbol{u} + (\nabla \cdot \boldsymbol{u})\boldsymbol{u} = (\boldsymbol{u} \cdot \nabla)\boldsymbol{u}. \qquad (5.5)
\end{aligned}
$$

- *Convective form and divergence form with modified pressure.* From (5.5), one obtains

$$
\begin{aligned}
(\boldsymbol{u} \cdot \nabla)\boldsymbol{u} + \nabla p &= \nabla \cdot \left(\boldsymbol{u}\boldsymbol{u}^T\right) + \frac{1}{2}\nabla\left(\boldsymbol{u}^T\boldsymbol{u}\right) - \frac{1}{2}\nabla\left(\boldsymbol{u}^T\boldsymbol{u}\right) + \nabla p \\
&= \nabla \cdot \left(\boldsymbol{u}\boldsymbol{u}^T\right) + \frac{1}{2}\left(\boldsymbol{u}^T\boldsymbol{u}\right) + \nabla p_{\text{mod}}
\end{aligned}
$$

  with

$$
p_{\text{mod}} = p - \frac{1}{2}\left(\boldsymbol{u}^T\boldsymbol{u}\right).
$$

- *Convective form and rotational form.* The rotational form goes also along with a redefinition of the pressure. In contrast to the equivalence of the convective form and the divergence form, it is not required that $\boldsymbol{u}$ is divergence-free. It holds

$$
\begin{aligned}
(\nabla \times \boldsymbol{u}) \times \boldsymbol{u} + \frac{1}{2}\nabla\left(\boldsymbol{u}^T\boldsymbol{u}\right) &= \begin{pmatrix} \partial_y u_3 - \partial_z u_2 \\ \partial_z u_1 - \partial_x u_3 \\ \partial_x u_2 - \partial_y u_1 \end{pmatrix} \times \boldsymbol{u} + \frac{1}{2}\nabla\left(u_1^2 + u_2^2 + u_3^2\right) \\
&= \begin{pmatrix} u_3\partial_z u_1 - u_3\partial_x u_3 - u_2\partial_x u_2 + u_2\partial_y u_1 \\ u_1\partial_x u_2 - u_1\partial_y u_1 - u_3\partial_y u_3 + u_3\partial_z u_2 \\ u_2\partial_y u_3 - u_2\partial_z u_2 - u_1\partial_z u_1 + u_1\partial_x u_3 \end{pmatrix} \\
&\quad + \frac{1}{2}\begin{pmatrix} 2u_1\partial_x u_1 + 2u_2\partial_x u_2 + 2u_3\partial_x u_3 \\ 2u_1\partial_y u_1 + 2u_2\partial_y u_2 + 2u_3\partial_y u_3 \\ 2u_1\partial_z u_1 + 2u_2\partial_z u_2 + 2u_3\partial_z u_3 \end{pmatrix} \\
&= \begin{pmatrix} u_1\partial_x u_1 + u_2\partial_y u_1 + u_3\partial_z u_1 \\ u_1\partial_x u_2 + u_2\partial_y u_2 + u_3\partial_z u_2 \\ u_1\partial_x u_3 + u_2\partial_y u_3 + u_3\partial_z u_3 \end{pmatrix} \\
&= (\boldsymbol{u} \cdot \nabla)\boldsymbol{u}. \qquad (5.6)
\end{aligned}
$$

  The term with the gradient is used to define a new pressure, the so-called Bernoulli pressure

$$
p_{\text{Bern}} = p + \frac{1}{2}\boldsymbol{u}^T\boldsymbol{u}. \qquad (5.7)
$$

  The derivation of identity (5.6) is also possible on the basis of (2.133).

□

Lemma 5.6. Basic properties of the convective term. *The convective term is trilinear, i.e., it is linear in each argument.*

Let $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w} \in H^1(\Omega)$, *then*

$$((\boldsymbol{u} \cdot \nabla)\, \boldsymbol{v}, \boldsymbol{w}) = \left(\nabla \cdot \left(\boldsymbol{v}\boldsymbol{u}^T\right), \boldsymbol{w}\right) - ((\nabla \cdot \boldsymbol{u})\, \boldsymbol{v}, \boldsymbol{w}), \tag{5.8}$$

$$((\boldsymbol{u} \cdot \nabla)\, \boldsymbol{v}, \boldsymbol{w}) = (\boldsymbol{u}, \nabla(\boldsymbol{v} \cdot \boldsymbol{w})) - ((\boldsymbol{u} \cdot \nabla)\, \boldsymbol{w}, \boldsymbol{v}). \tag{5.9}$$

Proof. The trilinearity of the convective term follows by the linearity of differentiation and integration, e.g., for $a, b \in \mathbb{R}$, one obtains

$$\begin{aligned}
&(((a\overline{\boldsymbol{u}} + b\hat{\boldsymbol{u}}) \cdot \nabla)\, \boldsymbol{v}, \boldsymbol{w}) \\
&= \int_\Omega \begin{pmatrix} (a\overline{u} + b\hat{u})_1 \partial_x v_1 + (a\overline{u} + b\hat{u})_2 \partial_y v_1 + (a\overline{u} + \hat{u})_3 \partial_z v_1 \\ (a\overline{u} + b\hat{u})_1 \partial_x v_2 + (a\overline{u} + b\hat{u})_2 \partial_y v_2 + (a\overline{u} + b\hat{u})_3 \partial_z v_2 \\ (a\overline{u} + b\hat{u})_1 \partial_x v_3 + (a\overline{u} + b\hat{u})_2 \partial_y v_3 + (a\overline{u} + b\hat{u})_3 \partial_z v_3 \end{pmatrix} \begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix} d\boldsymbol{x} \\
&= \int_\Omega a \begin{pmatrix} \overline{u}_1 \partial_x v_1 + \overline{u}_2 \partial_y v_1 + \overline{u}_3 \partial_z v_1 \\ \overline{u}_1 \partial_x v_2 + \overline{u}_2 \partial_y v_2 + \overline{u}_3 \partial_z v_2 \\ \overline{u}_1 \partial_x v_3 + \overline{u}_2 \partial_y v_3 + \overline{u}_3 \partial_z v_3 \end{pmatrix} d\boldsymbol{x} + \int_\Omega b \begin{pmatrix} \hat{u}_1 \partial_x v_1 + \hat{u}_2 \partial_y v_1 + \hat{u}_3 \partial_z v_1 \\ \hat{u}_1 \partial_x v_2 + \hat{u}_2 \partial_y v_2 + \hat{u}_3 \partial_z v_2 \\ \hat{u}_1 \partial_x v_3 + \hat{u}_2 \partial_y v_3 + \hat{u}_3 \partial_z v_3 \end{pmatrix} d\boldsymbol{x} \\
&= a\,((\overline{\boldsymbol{u}} \cdot \nabla)\, \boldsymbol{v}, \boldsymbol{w}) + b\,((\hat{\boldsymbol{u}} \cdot \nabla)\, \boldsymbol{v}, \boldsymbol{w}).
\end{aligned}$$

The calculations for the two other arguments of the convective term are similar.

For $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w} \in H^1(\Omega)$, the relations (5.8) and (5.9) are directly proved by straightforward calculations like in Remark 5.5. E.g., using the product rule and rearranging terms yields

$$\begin{aligned}
&(\boldsymbol{u}, \nabla(\boldsymbol{v} \cdot \boldsymbol{w})) \\
&= \int_\Omega \boldsymbol{u} \cdot \nabla (v_1 w_1 + v_2 w_2 + v_3 w_3)\ d\boldsymbol{x} \\
&= \int_\Omega \Big[u_1 v_1 \partial_x w_1 + u_1 v_2 \partial_x w_2 + u_1 v_3 \partial_x w_3 + u_2 v_1 \partial_y w_1 + u_1 v_2 \partial_y w_2 \\
&\quad + u_2 v_3 \partial_y w_3 + u_3 v_1 \partial_z w_1 + u_3 v_2 \partial_z w_2 + u_3 v_3 \partial_z w_3 \Big] \\
&\quad + \Big[u_1 w_1 \partial_x v_1 + u_1 w_2 \partial_x v_2 + u_1 w_3 \partial_x v_3 + u_2 w_1 \partial_y v_1 + u_1 w_2 \partial_y v_2 \\
&\quad + u_2 w_3 \partial_y v_3 + u_3 w_1 \partial_z v_1 + u_3 w_2 \partial_z v_2 + u_3 w_3 \partial_z v_3 \Big]\ d\boldsymbol{x} \\
&= ((\boldsymbol{u} \cdot \nabla)\, \boldsymbol{w}, \boldsymbol{v}) + ((\boldsymbol{u} \cdot \nabla)\, \boldsymbol{v}, \boldsymbol{w}).
\end{aligned}$$

■

Remark 5.7. Equivalent variational forms of the steady-state Navier–Stokes equations. In (5.2), the convective form of the nonlinear term

$$n_{\text{conv}}(\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}) = ((\boldsymbol{u} \cdot \nabla)\boldsymbol{v}, \boldsymbol{w})$$

was used. Equivalent variational forms are obtained by using the other forms of the nonlinear term described in Remark 5.5. The divergence form is defined by

$$n_{\text{div}}(\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}) = n_{\text{conv}}(\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}) + \frac{1}{2}\,((\nabla \cdot \boldsymbol{u})\, \boldsymbol{v}, \boldsymbol{w}).$$

The motivation for changing the factor in front of the divergence term in comparison with (5.5) is that the term will vanish if $\boldsymbol{v} = \boldsymbol{w}$ even if the first argument is not weakly divergence-free, see Lemma 5.9. This property is not true if a different factor than $1/2$ is used. If the rotational form

$$n_{\text{rot}}(\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}) = ((\nabla \times \boldsymbol{u}) \times \boldsymbol{v}, \boldsymbol{w})$$

is used, then the momentum equation in (5.2) changes to

$$(\nu \nabla \boldsymbol{u}, \nabla \boldsymbol{v}) + n_{\text{rot}}(\boldsymbol{u}, \boldsymbol{u}, \boldsymbol{v}) - (\nabla \cdot \boldsymbol{v}, p_{\text{Bern}}) = \langle \boldsymbol{f}, \boldsymbol{v} \rangle_{V',V} \quad \forall\, \boldsymbol{v} \in V,$$

where the Bernoulli pressure is defined in (5.7).

Finally, for the variational form of the Navier–Stokes equations, another form of the convective term can be applied. Integration by parts gives for $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w} \in H^1(\Omega)$

$$(\nabla \cdot \boldsymbol{u}, \boldsymbol{v} \cdot \boldsymbol{w}) = \int_\Gamma (\boldsymbol{v} \cdot \boldsymbol{w})(\boldsymbol{u} \cdot \boldsymbol{n})\, d\boldsymbol{s} - (\boldsymbol{u}, \nabla (\boldsymbol{v} \cdot \boldsymbol{w}))\,, \tag{5.10}$$

where $\boldsymbol{n}$ is the outward pointing unit normal vector on Γ. From the Sobolev imbeddings $H^1(\Omega) \to L^4(\Omega)$, see (A.16) and (A.17), it follows that $\boldsymbol{v}, \boldsymbol{w} \in L^4(\Omega)$ and consequently that $\boldsymbol{v} \cdot \boldsymbol{w} \in L^2(\Omega)$. Since Ω is a bounded domain, it follows that $\boldsymbol{v} \cdot \boldsymbol{w} \in L^1(\Omega)$. Then, there is a constant C such that $\boldsymbol{v} \cdot \boldsymbol{w} + C \in Q$. If $\boldsymbol{u}$ satisfies the second equation of (5.2), i.e., $\boldsymbol{u}$ is weakly divergence-free, and if $\boldsymbol{u} \cdot \boldsymbol{n} = 0$ on Γ, then it follows with integration by parts that

$$\begin{aligned} 0 = (\nabla \cdot \boldsymbol{u}, \boldsymbol{v} \cdot \boldsymbol{w} + C) &= (\nabla \cdot \boldsymbol{u}, \boldsymbol{v} \cdot \boldsymbol{w}) + \int_\Gamma C \boldsymbol{u} \cdot \boldsymbol{n}\, d\boldsymbol{s} - (\boldsymbol{u}, \nabla C) \\ &= (\nabla \cdot \boldsymbol{u}, \boldsymbol{v} \cdot \boldsymbol{w}). \end{aligned} \tag{5.11}$$

From (5.10), one obtains

$$(\boldsymbol{u}, \nabla (\boldsymbol{v} \cdot \boldsymbol{w})) = 0$$

and inserting this identity into (5.9) gives for $\boldsymbol{u}$ weakly divergence-free with $\boldsymbol{u} \cdot \boldsymbol{n} = 0$ on Γ and $\boldsymbol{v}, \boldsymbol{w} \in H^1(\Omega)$

$$n_{\text{conv}}(\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}) = -n_{\text{conv}}(\boldsymbol{u}, \boldsymbol{w}, \boldsymbol{v}). \tag{5.12}$$

With this relation, the skew-symmetric form of the convective term is defined by

$$n_{\text{skew}}(\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}) = \frac{1}{2} \left(n_{\text{conv}}(\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}) - n_{\text{conv}}(\boldsymbol{u}, \boldsymbol{w}, \boldsymbol{v}) \right).$$

□

Remark 5.8. Vanishing of the convective term. Analogously to the Oseen equations, the vanishing of the convective term if the last two arguments are identical is important for the analysis and finite element error analysis. For a discussion for cases when this property is given, it is referred to Remark 4.6. □

Lemma 5.9. Vanishing of the convective term. *Let* $\boldsymbol{u}, \boldsymbol{v} \in H^1(\Omega)$*, then*

$$n_{\text{rot}}(\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{v}) = n_{\text{skew}}(\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{v}) = 0. \tag{5.13}$$

If $\boldsymbol{u} \cdot \boldsymbol{n} = 0$ *or* $\boldsymbol{v} = \boldsymbol{0}$ *on* Γ*, then*

$$n_{\text{div}}(\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{v}) = 0. \tag{5.14}$$

If $\boldsymbol{u}$ *is weakly divergence-free and if* $\boldsymbol{u} \cdot \boldsymbol{n} = 0$ *on* Γ*, then*

$$n_{\text{conv}}(\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{v}) = 0. \tag{5.15}$$

Proof. The property (5.13) for $n_{\text{skew}}(\cdot,\cdot,\cdot)$ follows directly from the definition. A direct calculation gives for the rotational form

$$\begin{aligned}
&n_{\text{rot}}(\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{v}) \\
&= \left(\begin{pmatrix} v_3\partial_z u_1 - v_3\partial_x u_3 - v_2\partial_x u_2 + v_2\partial_y u_1 \\ v_1\partial_x u_2 - v_1\partial_y u_1 - v_3\partial_y u_3 + v_3\partial_z u_2 \\ v_2\partial_y u_3 - v_2\partial_z u_2 - v_1\partial_z u_1 + v_1\partial_x u_3 \end{pmatrix}, \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} \right) \\
&= \int_\Omega \Big[\partial_y u_1(v_1v_2 - v_1v_2) + \partial_z u_1(v_1v_3 - v_1v_3) + \partial_x u_2(v_1v_2 - v_1v_2) \\
&\quad + \partial_z u_2(v_2v_3 - v_2v_3) + \partial_x u_3(v_1v_3 - v_1v_3) + \partial_y u_3(v_2v_3 - v_2v_3) \Big] \, d\boldsymbol{x} \\
&= 0.
\end{aligned}$$

Considering the divergence form then one gets with (5.9) and (5.10)

$$\begin{aligned}
n_{\text{div}}(\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{v}) &= n_{\text{conv}}(\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{v}) + \frac{1}{2} \left((\nabla \cdot \boldsymbol{u}) \, \boldsymbol{v}, \boldsymbol{v} \right) \\
&= (\boldsymbol{u}, \nabla(\boldsymbol{v} \cdot \boldsymbol{v})) - n_{\text{conv}}(\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{v}) + \frac{1}{2} \left((\nabla \cdot \boldsymbol{u}) \, \boldsymbol{v}, \boldsymbol{v} \right) \\
&= -n_{\text{conv}}(\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{v}) - (\nabla \cdot \boldsymbol{u}, \boldsymbol{v} \cdot \boldsymbol{v}) + \frac{1}{2} \left((\nabla \cdot \boldsymbol{u}) \, \boldsymbol{v}, \boldsymbol{v} \right) \\
&= -n_{\text{div}}(\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{v}),
\end{aligned}$$

from what $n_{\text{div}}(\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{v}) = 0$ follows.

For the convective form, (5.15) follows from (5.12). Note that the derivation of (5.12) used the assumptions stated in the lemma for $n_{\text{conv}}(\cdot,\cdot,\cdot)$. ■

Lemma 5.10. Estimates of the convective term. *Let* $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w} \in H^1(\Omega)$*, where* Ω *is a bounded domain with Lipschitz boundary, then there is a* $C \in \mathbb{R}$ *such that*

$$|n_{\text{conv}}(\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w})| \le C \|\boldsymbol{u}\|_{H^1(\Omega)} \|\nabla \boldsymbol{v}\|_{L^2(\Omega)} \|\boldsymbol{w}\|_{H^1(\Omega)} . \tag{5.16}$$

For the skew-symmetric form, it holds

$$|n_{\text{skew}}(\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w})| \le C \|\boldsymbol{u}\|_{H^1(\Omega)} \|\boldsymbol{v}\|_{H^1(\Omega)} \|\boldsymbol{w}\|_{H^1(\Omega)} \tag{5.17}$$

and for the divergence form of the convective term

$$|n_{\mathrm{div}}(\boldsymbol{u},\boldsymbol{v},\boldsymbol{w})| \le C\,\|\boldsymbol{u}\|_{H^1(\Omega)}\,\|\boldsymbol{v}\|_{H^1(\Omega)}\,\|\boldsymbol{w}\|_{H^1(\Omega)}\,. \tag{5.18}$$

rotation form

Proof. The estimate starts with the application of the generalized Hölder inequality

$$|n_{\mathrm{conv}}(\boldsymbol{u},\boldsymbol{v},\boldsymbol{w})| = \left|\int_\Omega (\boldsymbol{u}\cdot\nabla)\,\boldsymbol{v}\cdot\boldsymbol{w}\;d\boldsymbol{x}\right| \le \|\boldsymbol{u}\|_{L^p(\Omega)}\,\|\nabla\boldsymbol{v}\|_{L^q(\Omega)}\,\|\boldsymbol{w}\|_{L^r(\Omega)}\,, \tag{5.19}$$

with

$$\frac{1}{p}+\frac{1}{q}+\frac{1}{r}=1,\;\; 1\le p,q,r\le\infty.$$

Since $\boldsymbol{v}\in H^1(\Omega)$, one can take at most $p=2$. The other two terms are of the same form such that they can treated similarly, i.e., one can take $p=r=4$. Applying the Sobolev imbedding $H^1(\Omega)\to L^4(\Omega)$, see (A.16) and (A.17), gives immediately the statement of the lemma for the convective term

$$|n_{\mathrm{conv}}(\boldsymbol{u},\boldsymbol{v},\boldsymbol{w})| \le C\,\|\boldsymbol{u}\|_{H^1(\Omega)}\,\|\nabla\boldsymbol{v}\|_{L^2(\Omega)}\,\|\boldsymbol{w}\|_{H^1(\Omega)}\,.$$

The statement for the skew-symmetric term follows by applying the triangle inequality

$$|n_{\mathrm{skew}}(\boldsymbol{u},\boldsymbol{v},\boldsymbol{w})| \le \frac{1}{2}\left(|n_{\mathrm{conv}}(\boldsymbol{u},\boldsymbol{v},\boldsymbol{w})| + |n_{\mathrm{conv}}(\boldsymbol{u},\boldsymbol{w},\boldsymbol{v})|\right)$$

and then estimate for the convective term.

For the divergence form of the convective term, one gets for the second part of this term, using the generalized Hölder's inequality, the Sobolev imbedding $H^1(\Omega)\to L^4(\Omega)$, and (2.38)

$$\begin{aligned}\frac{1}{2}\left((\nabla\cdot\boldsymbol{u})\,\boldsymbol{v},\boldsymbol{w}\right) &\le \frac{1}{2}\,\|\nabla\cdot\boldsymbol{u}\|_{L^2(\Omega)}\,\|\boldsymbol{v}\|_{L^4(\Omega)}\,\|\boldsymbol{w}\|_{L^4(\Omega)}\\ &\le C\,\|\nabla\cdot\boldsymbol{u}\|_{L^2(\Omega)}\,\|\boldsymbol{v}\|_{H^1(\Omega)}\,\|\boldsymbol{w}\|_{H^1(\Omega)}\\ &\le C\,\|\nabla\boldsymbol{u}\|_{L^2(\Omega)}\,\|\boldsymbol{v}\|_{H^1(\Omega)}\,\|\boldsymbol{w}\|_{H^1(\Omega)}\,.\end{aligned}$$

Since the first part of the divergence form is just the convective form, estimate (5.18) follows by combining the estimate for the second part and (5.16). ■

Remark 5.11. On the convective term.

- If $\boldsymbol{u},\boldsymbol{v},\boldsymbol{w}\in V$, then the application of the Poincaré–Friedrichs inequality (A.9) to (5.16) – (5.18) gives

$$|n_{\mathrm{conv}}(\boldsymbol{u},\boldsymbol{v},\boldsymbol{w})| \le C(\Omega)\,\|\boldsymbol{u}\|_V\,\|\boldsymbol{v}\|_V\,\|\boldsymbol{w}\|_V\,, \tag{5.20}$$
$$|n_{\mathrm{skew}}(\boldsymbol{u},\boldsymbol{v},\boldsymbol{w})| \le C(\Omega)\,\|\boldsymbol{u}\|_V\,\|\boldsymbol{v}\|_V\,\|\boldsymbol{w}\|_V\,, \tag{5.21}$$
$$|n_{\mathrm{div}}(\boldsymbol{u},\boldsymbol{v},\boldsymbol{w})| \le C(\Omega)\,\|\boldsymbol{u}\|_V\,\|\boldsymbol{v}\|_V\,\|\boldsymbol{w}\|_V\,. \tag{5.22}$$

- In the proof of Lemma 5.10, other choices of the parameters in Hölder's inequality are possible, which may lead to different estimates of the trilinear term. Yet other estimates can be derived for different regularity assumptions on the functions, e.g., see Layton and Tobiska (1998).

□

5.1.3 Existence, Uniqueness, and Stability of a Solution

Remark 5.12. Contents. This section presents results on the existence, uniqueness, and stability of a weak solution of the Navier–Stokes equations (5.2). It turns out that a solution always exists but it is unique only in the case of sufficiently small external forces and sufficiently large viscosity, see (5.26) below for the concrete requirement.

From the point of view of numerical simulations, the uniqueness case is the only interesting one. In the non-uniqueness case, small perturbations of the data will lead to time-dependent solutions. From the practical point of view, one should consider and discretize in such a case the time-dependent Navier–Stokes equations.

Theorem 5.13. Existence of a solution. *Let $\Omega \subset \mathbb{R}^d$, $d \in \{2,3\}$, be a bounded domain with Lipschitz boundary $\partial\Omega$ and let $\boldsymbol{f} \in H^{-1}(\Omega)$. Then there exists at least one solution of* (5.2).

Proof. For the proof it is referred to (Girault and Raviart, 1986, Chapter IV, Thm. 2.3 and 2.4). Essential ideas of the proof are as follows:

- The equivalent problem (5.4) in the divergence-free subspace V_{div} is considered, such that only the velocity appears.
- The problem is considered in finite-dimensional spaces (Galerkin method).
- A fixed point equation is constructed and the existence of a solution of the finite-dimensional problems is proved by the fixed point theorem of Brouwer[1].
- It is shown that for the dimension of the spaces going to infinity, a subsequence of the solutions tends to a solution of problem (5.4).
- The existence of the pressure is recovered with the help of the inf-sup condition (2.14).

■

Remark 5.14. Norms of the trilinear form of the convective term. The norm of the convective term is denoted by

$$N = \sup_{\boldsymbol{u},\boldsymbol{v},\boldsymbol{w}\in V\setminus\{\boldsymbol{0}\}} \frac{((\boldsymbol{u}\cdot\nabla)\boldsymbol{v},\boldsymbol{w})}{\|\boldsymbol{u}\|_V \|\boldsymbol{v}\|_V \|\boldsymbol{w}\|_V}.$$

Note that N is the smallest constant in estimate (5.20) since

$$N \geq \frac{((\boldsymbol{u}\cdot\nabla)\boldsymbol{v},\boldsymbol{w})}{\|\nabla\boldsymbol{u}\|_{L^2(\Omega)} \|\nabla\boldsymbol{v}\|_{L^2(\Omega)} \|\nabla\boldsymbol{w}\|_{L^2(\Omega)}} \quad \forall\, \boldsymbol{u},\boldsymbol{v},\boldsymbol{w}\in V,$$

which is the same inequality as (5.20). The existence of a smaller constant in (5.20) than N contradicts the definition of N. Likewise, define

[1] Brouwer's fixed point theorem. Let C denote a non-void, convex, and compact subset of a finite-dimensional space, and let F be continuous mapping from C into C. Then F has at least one fixed point.

$$N_{\mathrm{div}} = \sup_{\boldsymbol{u},\boldsymbol{v},\boldsymbol{w}\in V_{\mathrm{div}}\setminus\{\boldsymbol{0}\}} \frac{((\boldsymbol{u}\cdot\nabla)\boldsymbol{v},\boldsymbol{w})}{\|\boldsymbol{u}\|_V \|\boldsymbol{v}\|_V \|\boldsymbol{w}\|_V}. \qquad (5.23)$$

Since $V_{\mathrm{div}} \subset V$ and the supremum in a subset cannot be larger than the supremum in the whole set, if follows that $0 < N_{\mathrm{div}} \le N < \infty$. The boundedness of N is a consequence of estimate (5.20). □

Remark 5.15. Definition of an operator with the help of an Oseen problem. Let $\boldsymbol{b} \in V_{\mathrm{div}}$ and consider the Oseen problem: Find $\boldsymbol{u} \in V_{\mathrm{div}}$ such that

$$(\nu\nabla\boldsymbol{u},\nabla\boldsymbol{v}) + ((\boldsymbol{b}\cdot\nabla)\,\boldsymbol{u},\boldsymbol{v}) = \langle \boldsymbol{f},\boldsymbol{v}\rangle_{V',V} \quad \forall\, \boldsymbol{v}\in V_{\mathrm{div}}. \qquad (5.24)$$

This problem has a unique solution, see Theorem 4.7, and this solution satisfies the stability estimate

$$\|\nabla\boldsymbol{u}\|_{L^2(\Omega)} \le \frac{1}{\nu}\|\boldsymbol{f}\|_{H^{-1}(\Omega)}, \qquad (5.25)$$

see (4.7).

With the Oseen problem (5.24), an operator is defined which maps the given convection field $\boldsymbol{b}$ to the solution $\boldsymbol{u}$

$$N_{\mathrm{conv}} \;:\; V_{\mathrm{div}} \to V_{\mathrm{div}} \quad \boldsymbol{b}\mapsto\boldsymbol{u}.$$

It is obvious that each fixed point $\boldsymbol{u}_*$ of T is a velocity solution of the Navier–Stokes equations (5.4) since then if follows from (5.24) that

$$(\nu\nabla\boldsymbol{u}_*,\nabla\boldsymbol{v}) + ((\boldsymbol{u}_*\cdot\nabla)\,\boldsymbol{u}_*,\boldsymbol{v}) = \langle \boldsymbol{f},\boldsymbol{v}\rangle_{V',V} \quad \forall\, \boldsymbol{v}\in V_{\mathrm{div}}.$$

□

Theorem 5.16. Existence and uniqueness of a solution for small data. *Let the assumptions of Theorem 5.13 be satisfied and let in addition*

$$\frac{N_{\mathrm{div}}\|\boldsymbol{f}\|_{H^{-1}(\Omega)}}{\nu^2} < 1, \qquad (5.26)$$

then problem (5.4) *has a unique solution* $\boldsymbol{u}\in V_{\mathrm{div}}$ *and problem* (5.2) *has a unique solution* $(\boldsymbol{u},p)\in V\times Q$.

Proof. It will be shown that N_{conv} defines a contraction on V_{div}. First of all, it can be observed that N_{conv} is bounded independently of $\boldsymbol{b}$, since one obtains with (5.25)

$$\begin{aligned}\|N_{\mathrm{conv}}\| &= \sup_{\boldsymbol{b}\in V_{\mathrm{div}},\|\boldsymbol{b}\|_V=1}\|N_{\mathrm{conv}}\boldsymbol{b}\|_V = \sup_{\boldsymbol{b}\in V_{\mathrm{div}},\|\boldsymbol{b}\|_V=1}\|\boldsymbol{u}\|_V \\ &\le \sup_{\boldsymbol{b}\in V_{\mathrm{div}},\|\boldsymbol{b}\|_V=1}\frac{1}{\nu}\|\boldsymbol{f}\|_{H^{-1}(\Omega)} = \frac{1}{\nu}\|\boldsymbol{f}\|_{H^{-1}(\Omega)}.\end{aligned}$$

Now, one chooses $\boldsymbol{b}_1,\boldsymbol{b}_2\in V_{\mathrm{div}}$ arbitrary and denotes $\boldsymbol{u}_1 = N_{\mathrm{conv}}\boldsymbol{b}_1, \boldsymbol{u}_2 = N_{\mathrm{conv}}\boldsymbol{b}_2$. Both solutions $\boldsymbol{u}_1,\boldsymbol{u}_2$ are solution of the Oseen equation (5.24) with the same right-hand side. Subtracting these equations, one gets

$$\begin{aligned} 0 &= (\nu\nabla \boldsymbol{u}_1, \nabla \boldsymbol{v}) + ((\boldsymbol{b}_1 \cdot \nabla)\, \boldsymbol{u}_1, \boldsymbol{v}) - (\nu\nabla \boldsymbol{u}_2, \nabla \boldsymbol{v}) - ((\boldsymbol{b}_2 \cdot \nabla)\, \boldsymbol{u}_2, \boldsymbol{v}) \\ &= \nu\,(\nabla\,(\boldsymbol{u}_1 - \boldsymbol{u}_2)\,, \nabla \boldsymbol{v}) + (((\boldsymbol{b}_1 - \boldsymbol{b}_2)\cdot \nabla)\, \boldsymbol{u}_1, \boldsymbol{v}) + ((\boldsymbol{b}_2 \cdot \nabla)\,(\boldsymbol{u}_1 - \boldsymbol{u}_2)\,, \boldsymbol{v}) \quad \forall\, \boldsymbol{v} \in V_{\mathrm{div}}. \end{aligned}$$

Setting $\boldsymbol{v} = \boldsymbol{u}_1 - \boldsymbol{u}_2 \in V_{\mathrm{div}}$, the last term on the right-hand side vanishes because of (5.15) and one obtains with (5.23), (5.25), and (5.26)

$$\begin{aligned} \|\nabla\,(\boldsymbol{u}_1 - \boldsymbol{u}_2)\|_{L^2(\Omega)}^2 &= \|\boldsymbol{u}_1 - \boldsymbol{u}_2\|_V^2 = -\frac{1}{\nu}\,(((\boldsymbol{b}_1 - \boldsymbol{b}_2)\cdot\nabla)\,\boldsymbol{u}_1, \boldsymbol{u}_1 - \boldsymbol{u}_2) \\ &\le \frac{N_{\mathrm{div}}}{\nu}\, \|\boldsymbol{b}_1 - \boldsymbol{b}_2\|_V\, \|\boldsymbol{u}_1\|_V\, \|\boldsymbol{u}_1 - \boldsymbol{u}_2\|_V \\ &\le \frac{N_{\mathrm{div}}\, \|\boldsymbol{f}\|_{H^{-1}(\Omega)}}{\nu^2}\, \|\boldsymbol{b}_1 - \boldsymbol{b}_2\|_V\, \|\boldsymbol{u}_1 - \boldsymbol{u}_2\|_V \\ &< \|\boldsymbol{b}_1 - \boldsymbol{b}_2\|_V\, \|\boldsymbol{u}_1 - \boldsymbol{u}_2\|_V\,. \end{aligned}$$

It follows that

$$\|N_{\mathrm{conv}}\boldsymbol{b}_1 - N_{\mathrm{conv}}\boldsymbol{b}_2\|_V = \|\boldsymbol{u}_1 - \boldsymbol{u}_2\|_V < \|\boldsymbol{b}_1 - \boldsymbol{b}_2\|_V \quad \forall\, \boldsymbol{b}_1, \boldsymbol{b}_2 \in V_{\mathrm{div}},$$

which is the contraction property for N_{conv}. The existence and uniqueness of a solution of problem (5.4) follows now with the fixed point theorem of Banach, Theorem todo. The uniqueness of the solution of problem (5.2) is a consequence of the fact that V and Q satisfy the inf-sup condition (2.14), see Theorem 2.40. ■

Lemma 5.17. Stability of the solution. *Let* $(\boldsymbol{u}, p) \in V \times Q$ *be any solution of* (5.2)*, then*

$$\|\nabla \boldsymbol{u}\|_{L^2(\Omega)} \le \frac{1}{\nu}\, \|\boldsymbol{f}\|_{H^{-1}(\Omega)}\,, \tag{5.27}$$

$$\|p\|_{L^2(\Omega)} \le \frac{1}{\beta_{\mathrm{is}}} \left(2\, \|\boldsymbol{f}\|_{H^{-1}(\Omega)} + \frac{C}{\nu^2}\, \|\boldsymbol{f}\|_{H^{-1}(\Omega)}^2 \right). \tag{5.28}$$

Proof. The proof starts in the usual way by choosing as test function $(\boldsymbol{v}, q) = (\boldsymbol{u}, p)$ in (5.3)

$$(\nu\nabla \boldsymbol{u}, \nabla \boldsymbol{u}) + n(\boldsymbol{u}, \boldsymbol{u}, \boldsymbol{u}) = \langle \boldsymbol{f}, \boldsymbol{u}\rangle_{V',V}\,,$$

where $n(\cdot,\cdot,\cdot)$ is any of the convective terms introduced in Remark 5.7. With (5.13) – (5.15) it follows that

$$\nu\, \|\nabla \boldsymbol{u}\|_{L^2(\Omega)}^2 = \langle \boldsymbol{f}, \boldsymbol{u}\rangle_{V',V}.$$

The application of the inequality for the dual pairing gives

$$\nu\, \|\nabla \boldsymbol{u}\|_{L^2(\Omega)}^2 \le \|\boldsymbol{f}\|_{H^{-1}(\Omega)}\, \|\nabla \boldsymbol{u}\|_{L^2(\Omega)}\,.$$

This inequality is equivalent to (5.27).

Starting with the inf-sup condition, one obtains for the pressure with inserting (5.3), the estimate for the dual pairing, the Cauchy–Schwarz inequality (A.8), and (5.20)

$$\begin{aligned} \|p\|_{L^2(\Omega)} &\le \frac{1}{\beta_{\mathrm{is}}} \sup_{\boldsymbol{v}\in V\setminus\{\boldsymbol{0}\}} \frac{-(\nabla\cdot \boldsymbol{v}, p)}{\|\nabla \boldsymbol{v}\|_{L^2(\Omega)}} \\ &= \frac{1}{\beta_{\mathrm{is}}} \sup_{\boldsymbol{v}\in V\setminus\{\boldsymbol{0}\}} \frac{(\boldsymbol{f}, \boldsymbol{v}) - (\nu\nabla \boldsymbol{u}, \nabla \boldsymbol{v}) - n(\boldsymbol{u}, \boldsymbol{u}, \boldsymbol{v})}{\|\nabla \boldsymbol{v}\|_{L^2(\Omega)}} \\ &\le \frac{1}{\beta_{\mathrm{is}}} \Big(\|\boldsymbol{f}\|_{H^{-1}(\Omega)} + \nu\, \|\nabla \boldsymbol{u}\|_{L^2(\Omega)} + C\, \|\nabla \boldsymbol{u}\|_{L^2(\Omega)}^2 \Big). \end{aligned}$$

The statement follows now by inserting the stability estimate (5.27) for the velocity. ■

5.2 The Galerkin Finite Element Method

Remark 5.18. Contents. This section discusses finite element error estimates. The essential approaches were already presented for the Galerkin discretizations of the Stokes and the Oseen problem. For the steady-state Navier–Stokes equations, in addition an estimate of the nonlinear (trilinear) term is necessary. □

Remark 5.19. The Galerkin finite element formulation of the steady-state Navier–Stokes equations. Let $V^h \subset V$ and $Q^h \subset Q$ be inf-sup stable finite element spaces, i.e., (2.45) is fulfilled. Then, the Galerkin finite element discretization of the steady-state Navier–Stokes equations reads as follows: Find $(\boldsymbol{u}^h, p^h) \in V^h \times Q^h$ such that

$$\begin{aligned} \nu\left(\nabla \boldsymbol{u}^h, \nabla \boldsymbol{v}^h\right) + n\left(\boldsymbol{u}^h, \boldsymbol{u}^h, \boldsymbol{v}^h\right) - \left(\nabla \cdot \boldsymbol{v}^h, p^h\right) &= \langle \boldsymbol{f}, \boldsymbol{v}^h \rangle_{V',V} \ \forall\, \boldsymbol{v}^h \in V^h, \\ -\left(\nabla \cdot \boldsymbol{u}^h, q^h\right) &= 0 \qquad\qquad\quad \forall\, q^h \in Q^h, \end{aligned} \tag{5.29}$$

where $n(\cdot,\cdot,\cdot)$ is any of the convective terms introduced in Remark 5.7. An equivalent formulation is as follows: Find $(\boldsymbol{u}^h, p^h) \in V^h \times Q^h$ such that

$$\nu\left(\nabla \boldsymbol{u}^h, \nabla \boldsymbol{v}^h\right) + n\left(\boldsymbol{u}^h, \boldsymbol{u}^h, \boldsymbol{v}^h\right) - \left(\nabla \cdot \boldsymbol{v}^h, p^h\right) + \left(\nabla \cdot \boldsymbol{u}^h, q^h\right) = \langle \boldsymbol{f}, \boldsymbol{v}^h \rangle_{V',V} \tag{5.30}$$

for all $(\boldsymbol{v}^h, q^h) \in V^h \times Q^h$.

By the satisfaction of the discrete inf-sup condition, it is known that the space V^h_{div} is not empty, see Remark 2.16. reduced problem?

Existence and uniqueness of a solution are proved in the similar way as for the continuous equation. In particular, a unique solution can be expected only for small external forces (right-hand side) and a large viscosity. A discrete analog to condition (5.26) will be assumed throughout this section. □

Remark 5.20. On the convective term for the finite element error analysis. A main tool in the analysis of the Oseen equations is estimate (4.4) which states that the convective term vanishes if the convection field is (weakly) divergence-free and the ansatz and test functions are identical. A similar property for the different types of the nonlinear convective term of the continuous Navier–Stokes equations was proved in Lemma 5.9.

For $n_{\mathrm{conv}}(\cdot,\cdot,\cdot)$, the proof of this property relies on the assumption that the convection field is weakly divergence-free. Since generally $V^h_{\mathrm{div}} \not\subset V_{\mathrm{div}}$, see Remark 2.47, finite element velocity fields will be generally not weakly divergence-free and the convective term $n_{\mathrm{conv}}\left(\boldsymbol{u}^h, \boldsymbol{v}^h, \boldsymbol{v}^h\right)$ will not vanish for $\boldsymbol{u}^h \in V^h_{\mathrm{div}}$ and $\boldsymbol{v}^h \in V^h$.

In contrast, it is obvious to observe that

$$n_{\mathrm{skew}}\left(\boldsymbol{u}^h,\boldsymbol{v}^h,\boldsymbol{v}^h\right) = \frac{1}{2}\left(n_{\mathrm{conv}}\left(\boldsymbol{u}^h,\boldsymbol{v}^h,\boldsymbol{v}^h\right) - n_{\mathrm{conv}}\left(\boldsymbol{u}^h,\boldsymbol{v}^h,\boldsymbol{v}^h\right)\right) = 0 \quad (5.31)$$

for all $\boldsymbol{u}^h, \boldsymbol{v}^h \in V^h$. Analogously to the proof of Lemma 5.9, one finds by direct computations that also

$$n_{\mathrm{rot}}\left(\boldsymbol{u}^h,\boldsymbol{v}^h,\boldsymbol{v}^h\right) = 0 \quad \forall\, \boldsymbol{u}^h, \boldsymbol{v}^h \in V^h \quad (5.32)$$

and

$$n_{\mathrm{div}}\left(\boldsymbol{u}^h,\boldsymbol{v}^h,\boldsymbol{v}^h\right) = 0 \quad \forall\, \boldsymbol{u}^h, \boldsymbol{v}^h \in V^h. \quad (5.33)$$

Main tools in the finite element error analysis of the convective term are properties of the form (5.31) – (5.33), and estimates of the convective term from above, see (5.21) and (5.22).

Usually, one finds in the literature the analysis carried out for the skew-symmetric form of the convective term. The presentation below will also use this form. □

Theorem 5.21. Existence and uniqueness of the Galerkin finite element solution for small data. *Let the assumptions of Theorem 5.13 be satisfied and let in addition*

$$\frac{N_{\mathrm{div}}^h \left\|\boldsymbol{f}\right\|_{H^{-1}(\Omega)}}{\nu^2} < 1, \quad (5.34)$$

where

$$N_{\mathrm{div}}^h = \sup_{\boldsymbol{u}^h,\boldsymbol{v}^h,\boldsymbol{w}^h \in V_{\mathrm{div}}^h \setminus \{\boldsymbol{0}\}} \frac{\frac{1}{2}\left(((\boldsymbol{u}^h\cdot\nabla)\boldsymbol{v}^h,\boldsymbol{w}^h) - ((\boldsymbol{u}^h\cdot\nabla)\boldsymbol{v}^h,\boldsymbol{w}^h)\right)}{\|\boldsymbol{u}^h\|_V\,\|\boldsymbol{v}^h\|_V\,\|\boldsymbol{w}^h\|_V}. \quad (5.35)$$

If a pair of finite element spaces $V^h \times Q^h$ is used which satisfies the discrete inf-sup condition (2.45)*, then problem* (5.29) *with $n\left(\boldsymbol{u}^h,\boldsymbol{u}^h,\boldsymbol{v}^h\right) = n_{\mathrm{skew}}\left(\boldsymbol{u}^h,\boldsymbol{u}^h,\boldsymbol{v}^h\right)$ has a unique solution $(\boldsymbol{u}^h,p^h) \in V^h \times Q^h$.*

Proof. The proof follows the lines for proving the uniqueness of the solution for the continuous problem, see Theorem 5.16. One considers now an Oseen problem in V_{div}^h: Given $\boldsymbol{b}^h \in V_{\mathrm{div}}^h$, find $\boldsymbol{u}^h \in V_{\mathrm{div}}^h$ such that for given $\boldsymbol{f} \in V'$

$$\left(\nu\nabla\boldsymbol{u}^h,\nabla\boldsymbol{v}^h\right) + \frac{1}{2}\left(\left(\left(\boldsymbol{b}^h\cdot\nabla\right)\boldsymbol{u}^h,\boldsymbol{v}^h\right) - \left(\left(\boldsymbol{b}^h\cdot\nabla\right)\boldsymbol{v}^h,\boldsymbol{u}^h\right)\right) = \langle\boldsymbol{f},\boldsymbol{v}\rangle_{V',V} \quad \forall\, \boldsymbol{v} \in V_{\mathrm{div}}^h.$$

Analogously to Corallary 4.12 the existence of a unique solution of this problem is proved and a stability estimate of the form (4.12) is derived as in Lemma 4.13. Then, one defines a linear operator $N_{\mathrm{conv}}^h : V_{\mathrm{div}}^h \to V_{\mathrm{div}}^h$ which maps $\boldsymbol{b}^h \to \boldsymbol{u}^h$. Now, one shows, analogously to the proof of Theorem 5.16, that N_{conv}^h is bounded and it is a contraction. ■

Lemma 5.22. Stability of the finite element solution. *Let $V^h \times Q^h$ be an inf-sup stable finite element spaces. Then, the finite element solution of the steady-state Navier–Stokes equations with skew-symmetric form of the convective term is stable*

$$\left\|\nabla \boldsymbol{u}^h\right\|_{L^2(\Omega)} \le \frac{1}{\nu} \left\|\boldsymbol{f}\right\|_{H^{-1}(\Omega)}, \tag{5.36}$$

$$\left\|p^h\right\|_{L^2(\Omega)} \le \frac{1}{\beta_{\mathrm{is}}^h} \left(2 \left\|\boldsymbol{f}\right\|_{H^{-1}(\Omega)} + \frac{C}{\nu^2} \left\|\boldsymbol{f}\right\|_{H^{-1}(\Omega)}^2\right). \tag{5.37}$$

Proof. The proof is performed analogously to the proof of Lemma 5.17. ■

Theorem 5.23. Finite element error estimate for the $L^2(\Omega)$ norm of the gradient of the velocity. *Let $\Omega \subset \mathbb{R}^d$, $d \in \{2,3\}$, be a bounded domain with polyhedral and Lipschitz-continuous boundary, let* (5.26) *and let instead of* (5.34) *the stronger condition*

$$\frac{N_{\mathrm{div}}^h \left\|\boldsymbol{f}\right\|_{H^{-1}(\Omega)}}{\nu^2} \le \frac{1}{4}, \tag{5.38}$$

be satisfied. Let $(\boldsymbol{u}, p) \in V \times Q$ be the unique solution of the Navier–Stokes equations (5.2). *Assume that this problem is discretized with inf-sup stable finite element spaces $V^h \times Q^h$ using the skew-symmetric form of the convective term and denote by $\boldsymbol{u}^h \in V_{\mathrm{div}}^h$ the velocity solution. Then, the following error estimate holds*

$$\begin{aligned}&\left\|\nabla(\boldsymbol{u} - \boldsymbol{u}^h)\right\|_{L^2(\Omega)} \\ &\le C\left(\left(1 + \frac{1}{\nu^2}\left\|\boldsymbol{f}\right\|_{H^{-1}(\Omega)}\right) \inf_{\boldsymbol{v}^h \in V_{\mathrm{div}}^h} \left\|\nabla\left(\boldsymbol{u} - \boldsymbol{v}^h\right)\right\|_{L^2(\Omega)} \right. \\ &\qquad \left. + \frac{1}{\nu} \inf_{q^h \in Q^h} \left\|p - q^h\right\|_{L^2(\Omega)}\right).\end{aligned} \tag{5.39}$$

The constant C does not depend on the mesh size.

Proof. The principle of the proof is the same as for the Stokes equations, see Theorem 3.20. Since the space V_{div}^h is not empty, one can use test functions $\boldsymbol{v}^h \in V_{\mathrm{div}}^h$ in (5.3) and (5.30) and subtract these equations to obtain the following error equation

$$\nu\left(\nabla(\boldsymbol{u} - \boldsymbol{u}^h), \nabla \boldsymbol{v}^h\right) + n_{\mathrm{skew}}\left(\boldsymbol{u}, \boldsymbol{u}, \boldsymbol{v}^h\right) - n_{\mathrm{skew}}\left(\boldsymbol{u}^h, \boldsymbol{u}^h, \boldsymbol{v}^h\right) - \left(\nabla \cdot \boldsymbol{v}^h, p - q^h\right) = 0$$

for all $\boldsymbol{v}^h \in V_{\mathrm{div}}^h$ and all $q^h \in Q^h$. In this step, also $\left(\nabla \cdot \boldsymbol{v}^h, q^h\right)$ for all $q^h \in Q^h$ was applied. Next, the error is decomposed in an approximation error and a discrete remainder

$$\boldsymbol{u} - \boldsymbol{u}^h = \left(\boldsymbol{u} - I^h \boldsymbol{u}\right) - \left(\boldsymbol{u}^h - I^h \boldsymbol{u}\right) = \boldsymbol{\eta} - \boldsymbol{\phi}^h, \quad I^h \boldsymbol{u} \in V_{\mathrm{div}}^h. \tag{5.40}$$

Inserting this decomposition into the error equation and setting $\boldsymbol{v}^h = \boldsymbol{\phi}^h$ leads to

$$\begin{aligned}\nu\left\|\nabla \boldsymbol{\phi}^h\right\|_{L^2(\Omega)}^2 &= \nu\left(\nabla \boldsymbol{\eta}, \nabla \boldsymbol{\phi}^h\right) - \left(\nabla \cdot \boldsymbol{\phi}^h, p - q^h\right) \\ &\quad + n_{\mathrm{skew}}\left(\boldsymbol{u}, \boldsymbol{u}, \boldsymbol{\phi}^h\right) - n_{\mathrm{skew}}\left(\boldsymbol{u}^h, \boldsymbol{u}^h, \boldsymbol{\phi}^h\right), \quad \forall\, q^h \in Q^h.\end{aligned} \tag{5.41}$$

The first two terms are estimated in a similar way as for the Oseen equations, see the proof of Theorem 4.14,

$$\nu\left(\nabla \boldsymbol{\eta}, \nabla \boldsymbol{\phi}^h\right) \le 2\nu \left\|\nabla \boldsymbol{\eta}\right\|_{L^2(\Omega)}^2 + \frac{\nu}{8}\left\|\nabla \boldsymbol{\phi}^h\right\|_{L^2(\Omega)}^2,$$

and

$$\left(\nabla \cdot \boldsymbol{\phi}^h, p - q^h\right) \le \frac{2}{\nu} \left\|p - q^h\right\|_{L^2(\Omega)}^2 + \frac{\nu}{8} \left\|\nabla \boldsymbol{\phi}^h\right\|_{L^2(\Omega)}^2 .$$

The new aspect for the Navier–Stokes equations is the estimate of the trilinear terms. Such terms are written in the form

$$aa - bb = aa - ab + ab - bb = a(a-b) + (a-b)b.$$

The differences are used to introduce approximation errors into the estimate. Applying this approach, using (5.40), yields

$$\begin{aligned}
&n_{\text{skew}}\left(\boldsymbol{u}, \boldsymbol{u}, \boldsymbol{\phi}^h\right) - n_{\text{skew}}\left(\boldsymbol{u}^h, \boldsymbol{u}^h, \boldsymbol{\phi}^h\right) \\
&= n_{\text{skew}}\left(\boldsymbol{u}, \boldsymbol{u}, \boldsymbol{\phi}^h\right) - n_{\text{skew}}\left(\boldsymbol{u}, \boldsymbol{u}^h, \boldsymbol{\phi}^h\right) + n_{\text{skew}}\left(\boldsymbol{u}, \boldsymbol{u}^h, \boldsymbol{\phi}^h\right) - n_{\text{skew}}\left(\boldsymbol{u}^h, \boldsymbol{u}^h, \boldsymbol{\phi}^h\right) \\
&= n_{\text{skew}}\left(\boldsymbol{u}, \boldsymbol{u} - \boldsymbol{u}^h, \boldsymbol{\phi}^h\right) + n_{\text{skew}}\left(\boldsymbol{u} - \boldsymbol{u}^h, \boldsymbol{u}^h \boldsymbol{\phi}^h\right) \qquad (5.42) \\
&= n_{\text{skew}}\left(\boldsymbol{u}, \boldsymbol{\eta}, \boldsymbol{\phi}^h\right) - n_{\text{skew}}\left(\boldsymbol{u}, \boldsymbol{\phi}^h, \boldsymbol{\phi}^h\right) + n_{\text{skew}}\left(\boldsymbol{\eta}, \boldsymbol{u}^h, \boldsymbol{\phi}^h\right) - n_{\text{skew}}\left(\boldsymbol{\phi}^h, \boldsymbol{u}^h, \boldsymbol{\phi}^h\right).
\end{aligned}$$

The term with $\boldsymbol{\phi}^h$ in the last two arguments vanishes by (5.31). Now, all terms are estimated separately, using (5.21), Young's inequality (A.4), and the stability estimates (5.27) and (5.36). For the first term, one obtains

$$\begin{aligned}
n_{\text{skew}}\left(\boldsymbol{u}, \boldsymbol{\eta}, \boldsymbol{\phi}^h\right) &\le C \left\|\nabla \boldsymbol{u}\right\|_{L^2(\Omega)} \left\|\nabla \boldsymbol{\eta}\right\|_{L^2(\Omega)} \left\|\nabla \boldsymbol{\phi}^h\right\|_{L^2(\Omega)} \\
&\le \frac{2C}{\nu} \left\|\nabla \boldsymbol{\eta}\right\|_{L^2(\Omega)}^2 \left\|\nabla \boldsymbol{u}\right\|_{L^2(\Omega)}^2 + \frac{\nu}{8} \left\|\nabla \boldsymbol{\phi}^h\right\|_{L^2(\Omega)}^2 \\
&\le \frac{C}{\nu^3} \left\|\nabla \boldsymbol{\eta}\right\|_{L^2(\Omega)}^2 \left\|\boldsymbol{f}\right\|_{H^{-1}(\Omega)}^2 + \frac{\nu}{8} \left\|\nabla \boldsymbol{\phi}^h\right\|_{L^2(\Omega)}^2 .
\end{aligned}$$

The estimate for the third term is performed analogously, yielding

$$n_{\text{skew}}\left(\boldsymbol{\eta}, \boldsymbol{u}^h, \boldsymbol{\phi}^h\right) \le \frac{C}{\nu^3} \left\|\nabla \boldsymbol{\eta}\right\|_{L^2(\Omega)}^2 \left\|\boldsymbol{f}\right\|_{H^{-1}(\Omega)}^2 + \frac{\nu}{8} \left\|\nabla \boldsymbol{\phi}^h\right\|_{L^2(\Omega)}^2 .$$

The problematic term, for which the assumption on the smallness of the data is required, is the last one. With (5.35), (5.36), and (5.38), one gets

$$\begin{aligned}
n_{\text{skew}}\left(\boldsymbol{\phi}^h, \boldsymbol{u}^h, \boldsymbol{\phi}^h\right) &\le N_{\text{div}}^h \left\|\nabla \boldsymbol{u}^h\right\|_{L^2(\Omega)} \left\|\nabla \boldsymbol{\phi}^h\right\|_{L^2(\Omega)}^2 \\
&\le \frac{4 N_{\text{div}}^h \left\|\boldsymbol{f}\right\|_{H^{-1}(\Omega)}}{\nu^2} \left(\frac{\nu}{4} \left\|\nabla \boldsymbol{\phi}^h\right\|_{L^2(\Omega)}^2\right) \\
&\le \frac{\nu}{4} \left\|\nabla \boldsymbol{\phi}^h\right\|_{L^2(\Omega)}^2 .
\end{aligned}$$

This term can be absorbed into the left-hand side of (5.41).

Substituting all estimates into (5.41) gives

$$\begin{aligned}
&\left\|\nabla \boldsymbol{\phi}^h\right\|_{L^2(\Omega)}^2 \\
&\le C \left(\left\|\nabla \boldsymbol{\eta}\right\|_{L^2(\Omega)}^2 + \frac{1}{\nu^2} \left\|p - q^h\right\|_{L^2(\Omega)}^2 + \frac{1}{\nu^4} \left\|\nabla \boldsymbol{\eta}\right\|_{L^2(\Omega)}^2 \left\|\boldsymbol{f}\right\|_{H^{-1}(\Omega)}^2\right).
\end{aligned}$$

The application of the triangle inequality finally leads to

$$\begin{aligned}
&\left\|\nabla\left(\boldsymbol{u}-\boldsymbol{u}^h\right)\right\|_{L^2(\Omega)} \\
&\le \|\nabla\boldsymbol{\eta}\|_{L^2(\Omega)} + \left\|\nabla\boldsymbol{\phi}^h\right\|_{L^2(\Omega)} \\
&\le C\left(\left(1+\frac{1}{\nu^2}\|\boldsymbol{f}\|_{H^{-1}(\Omega)}\right)\|\nabla\boldsymbol{\eta}\|_{L^2(\Omega)} + \frac{1}{\nu}\left\|p-q^h\right\|_{L^2(\Omega)}\right).
\end{aligned}$$

This estimate gives the statement of the theorem. ■

Remark 5.24. On Theorem 5.23. From the proof of Theorem 5.23 it is clear, that condition (5.38) can be relaxed to

$$\frac{N_{\mathrm{div}}^h \|\boldsymbol{f}\|_{H^{-1}(\Omega)}}{\nu^2} \le q < 1$$

by applying different scalings in the applications of Young's inequality. However, large values of q in the analysis lead to a large constant C in the error estimate (5.39). □

Theorem 5.25. Finite element error estimate for the $L^2(\Omega)$ norm of the pressure. *Let the assumptions of Theorem 5.23 be fulfilled, then the following error estimate for the pressure holds*

$$\begin{aligned}
&\left\|p-p^h\right\|_{L^2(\Omega)} \\
&\le C\frac{\nu}{\beta_{\mathrm{is}}^h}\left(1+\frac{\|\boldsymbol{f}\|_{H^{-1}(\Omega)}}{\nu^2}\right)^2 \inf_{\boldsymbol{v}^h\in V_{\mathrm{div}}^h}\left\|\nabla\left(\boldsymbol{u}-\boldsymbol{v}^h\right)\right\|_{L^2(\Omega)} \\
&\quad + C\left(1+\frac{1}{\beta_{\mathrm{is}}^h}\left(1+\frac{\|\boldsymbol{f}\|_{H^{-1}(\Omega)}}{\nu^2}\right)\right)\inf_{q^h\in Q^h}\left\|p-q^h\right\|_{L^2(\Omega)}.
\end{aligned} \tag{5.43}$$

The constants do not depend on the mesh size.

Proof. The proof follows the lines of the proofs of Theorems 3.24 and 4.15.

The triangle inequality gives for all $q^h \in Q^h$

$$\left\|p-p^h\right\|_{L^2(\Omega)} \le \left\|p-q^h\right\|_{L^2(\Omega)} + \left\|p^h-q^h\right\|_{L^2(\Omega)}. \tag{5.44}$$

The estimate of the second term starts with the discrete inf-sup condition (2.45) and the insertion of the finite element problem (5.29) as well as the variational form of the steady-state Navier–Stokes equations (5.2)

$$\begin{aligned}
&\left\|p^h-q^h\right\|_{L^2(\Omega)} \\
&\le \frac{1}{\beta_{\mathrm{is}}^h}\sup_{\boldsymbol{v}^h\in V^h}\frac{b\left(\boldsymbol{v}^h,p^h-q^h\right)}{\left\|\nabla\boldsymbol{v}^h\right\|_{L^2(\Omega)}} \\
&= \frac{1}{\beta_{\mathrm{is}}^h}\sup_{\boldsymbol{v}^h\in V^h}\left(\frac{1}{\left\|\nabla\boldsymbol{v}^h\right\|_{L^2(\Omega)}}\right. \\
&\quad\left.\times\left(\nu\left(\nabla(\boldsymbol{u}-\boldsymbol{u}^h),\nabla\boldsymbol{v}^h\right) + n_{\mathrm{skew}}\left(\boldsymbol{u},\boldsymbol{u},\boldsymbol{v}^h\right) - n_{\mathrm{skew}}\left(\boldsymbol{u}^h,\boldsymbol{u}^h,\boldsymbol{v}^h\right) - \left(\nabla\cdot\boldsymbol{v}^h,p-q^h\right)\right)\right).
\end{aligned}$$

In the next step, an identity like (5.42) is used

$$\begin{aligned} & n_{\text{skew}}\left(\boldsymbol{u}, \boldsymbol{u}, \boldsymbol{v}^h\right) - n_{\text{skew}}\left(\boldsymbol{u}^h, \boldsymbol{u}^h, \boldsymbol{v}^h\right) \\ & = n_{\text{skew}}\left(\boldsymbol{u}, \boldsymbol{u} - \boldsymbol{u}^h, \boldsymbol{v}^h\right) + n_{\text{skew}}\left(\boldsymbol{u} - \boldsymbol{u}^h, \boldsymbol{u}^h, \boldsymbol{v}^h\right). \end{aligned}$$

Inserting this identity, using the Cauchy–Schwarz inequality (A.8), the estimate (5.21) for the trilinear term, (2.144), and the stability estimates (5.27) and (5.36) leads to

$$\begin{aligned} & \left\|p^h - q^h\right\|_{L^2(\Omega)} \\ & \le \frac{1}{\beta_{\text{is}}^h}\Big(\nu\left\|\nabla(\boldsymbol{u}-\boldsymbol{u}^h)\right\|_{L^2(\Omega)} + C\left\|\nabla\boldsymbol{u}\right\|_{L^2(\Omega)}\left\|\nabla(\boldsymbol{u}-\boldsymbol{u}^h)\right\|_{L^2(\Omega)} \\ & \quad + C\left\|\nabla\boldsymbol{u}^h\right\|_{L^2(\Omega)}\left\|\nabla(\boldsymbol{u}-\boldsymbol{u}^h)\right\|_{L^2(\Omega)} + \left\|p - q^h\right\|_{L^2(\Omega)}\Big) \\ & = \frac{C}{\beta_{\text{is}}^h}\left(\left(\nu + \frac{\|\boldsymbol{f}\|_{H^{-1}(\Omega)}}{\nu}\right)\left\|\nabla(\boldsymbol{u}-\boldsymbol{u}^h)\right\|_{L^2(\Omega)} + \left\|p - q^h\right\|_{L^2(\Omega)}\right). \end{aligned}$$

Using this estimate in (5.44) gives

$$\begin{aligned} & \left\|p - p^h\right\|_{L^2(\Omega)} \\ & \le \frac{C}{\beta_{\text{is}}^h}\nu\left(1 + \frac{\|\boldsymbol{f}\|_{H^{-1}(\Omega)}}{\nu^2}\right)\left\|\nabla(\boldsymbol{u}-\boldsymbol{u}^h)\right\|_{L^2(\Omega)} + \left(1 + \frac{C}{\beta_{\text{is}}^h}\right)\left\|p - q^h\right\|_{L^2(\Omega)}. \end{aligned}$$

Inserting now estimate (5.39) for $\left\|\nabla(\boldsymbol{u}-\boldsymbol{u}^h)\right\|_{L^2(\Omega)}$ finishes the proof. ■

L^2 error velocity ?

Corollary 5.26. Finite element error estimates for conforming pairs of finite element spaces. *Let the assumptions of Theorem 5.23 be fulfilled, let $\Omega \subset \mathbb{R}^d$, $d \in \{2,3\}$, be a bounded domain with polyhedral and Lipschitz-continuous boundary which is decomposed by a regular and quasi-uniform family of triangulations $\{\mathcal{T}^h\}$, and let $(\boldsymbol{u}, p)$ be the unique solution of the steady-state Navier–Stokes equations equations (5.2) with $\boldsymbol{u} \in H^{k+1}(\Omega) \cap V$ and $p \in H^k(\Omega) \cap Q$. Then for the inf-sup stable pairs of finite element spaces*

- P_k^{bubble}/P_k, $k = 1$ *(MINI element),*
- P_k/P_{k-1}, Q_k/Q_{k-1}, $k \ge 2$ *(Taylor–Hood element),*
- $P_2^{\text{bubble}}/P_1^{\text{disc}}$, $P_2^{\text{BR}}/P_1^{\text{disc}}$, $Q_k/P_{k-1}^{\text{disc}}$, $k \ge 2$,

the following error estimates hold

$$\left\|\nabla(\boldsymbol{u}-\boldsymbol{u}^h)\right\|_{L^2(\Omega)} \le Ch^k\left(\frac{1}{\nu^2}\|\boldsymbol{u}\|_{H^{k+1}(\Omega)} + \frac{1}{\nu}\|p\|_{H^k(\Omega)}\right), \quad (5.45)$$

$$\left\|p - p^h\right\|_{L^2(\Omega)} \le Ch^k\left(\frac{1}{\nu^3}\|\boldsymbol{u}\|_{H^{k+1}(\Omega)} + \frac{1}{\nu^2}\|p\|_{H^k(\Omega)}\right). \quad (5.46)$$

The constant in (5.45) might depend and the constant in (5.46) does depend on the inverse of the discrete inf-sup constant β_{is}^h.

Proof. The estimates follow directly from (5.39), (5.43), and the approximation properties of the finite element spaces. Either there is an interpolant in V_{div}^h with optimal approximation properties or one can apply Lemma 2.51. In the latter case, the constant in (5.45) depends on $\left(\beta_{\mathrm{is}}^h\right)^{-1}$. The dependency of the constant in (5.46) on $\left(\beta_{\mathrm{is}}^h\right)^{-1}$ follows directly from (5.43). ■

Remark 5.27. Discussion of dependency on ν, comparison with the Oseen equations. For the Oseen equations, the constants in the error bounds for the $L^2(\Omega)$ errors for the gradient of the velocity and the pressure depend only on ν^{-1}, compare the discussion in Remark 4.17. It follows from (5.45) and (5.43) that there is a stronger dependency for the Navier–Stokes equations. Formally, the difference arises on the one hand from absorbing the term $\|\boldsymbol{b}\|_{L^\infty(\Omega)}$ from the convection field into the constants of the error estimates for the Oseen equations because $\|\boldsymbol{b}\|_{L^\infty(\Omega)} = \mathcal{O}(1)$ is assumed. On the other hand, the terms $\|\nabla \boldsymbol{u}\|_{L^2(\Omega)}$ and $\left\|\nabla \boldsymbol{u}^h\right\|_{L^2(\Omega)}$ are estimated with the stability bounds (5.27) and (5.36) in the case of the Navier–Stokes equations, which gives an extra negative power of ν. Assuming that the convection field of the Oseen equations behaves like the convection field of the Navier–Stokes equations, then the constant C_{os} from (4.14) scales like $\mathcal{O}\left(\nu^{-3/2}\right)$ such that one obtains in the error estimates (4.21) and (4.22) (for $c = 0$) the same dependency on the viscosity as for the Navier–Stokes equations. □

Example 5.28. Analytical example which supports the error estimates. Example D.2 is considered, for detailed information to the simulations it is also referred to Example 3.29.

Figures 5.1 – 5.3 show results which were obtained for $\nu = 10^{-2}$ and the regular grids from Figure 3.2. The stopping criterion for the solution of the nonlinear problems was the requirement that the Euclidean norm of the residual vector was less than 10^{-10}. The convective form of the convective term was used in the simulations.

The order of convergence for errors in different norms coincide generally with the predictions from the numerical analysis. Only for the $L^2(\Omega)$ norm of the pressure and the mini element $P_1^{\mathrm{bubble}}/P_1$, a higher order than expected can be observed. Since the solution of this example is from Q_4/Q_3, it is reproduced on all grids if this pair of finite element spaces is used.

Figure 5.4 presents results for the Q_2/Q_1 finite element and different values of ν. For $\nu = 10^{-3}$, no way was found to solve the nonlinear problem on the coarsest grids. One can observe on coarser grids larger velocity errors for smaller coefficients ν.

ausführlicher ? □

Example 5.29. Flow around a cylinder in two dimensions. This problem is described in Example D.4.

Simulations were performed with the convective form $n_{\mathrm{conv}}(\cdot,\cdot,\cdot)$ of the convective term and different inf-sup stable pairs of finite element spaces on triangular and quadrilateral grids. The initial grids (level 0) are presented in Figure 5.5. The considered problem, with Dirichlet conditions at the outflow (D.4) was studied comprehensively in John and Matthies (2001). In these studies, it was shown that the use of isoparametric finite elements at the

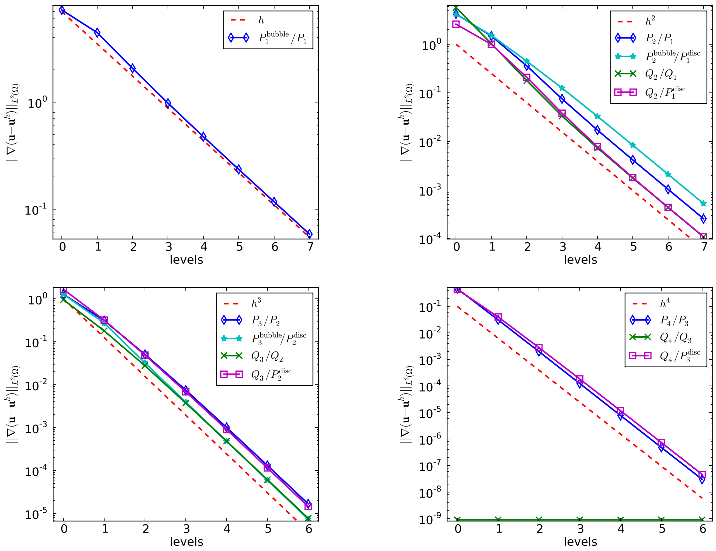

Fig. 5.1 Example 5.28. Convergence of the errors $\left\|\nabla(\boldsymbol{u}-\boldsymbol{u}^h)\right\|_{L^2(\Omega)}$ for different discretizations with different orders k.

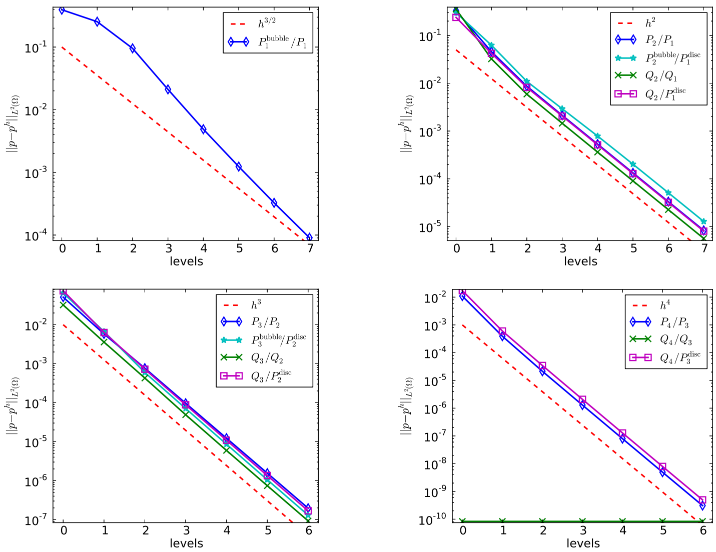

Fig. 5.2 Example 5.28. Convergence of the errors $\left\|p-p^h\right\|_{L^2(\Omega)}$ for different discretizations with different orders k.

cylinder was essential for obtaining accurate results for higher order finite

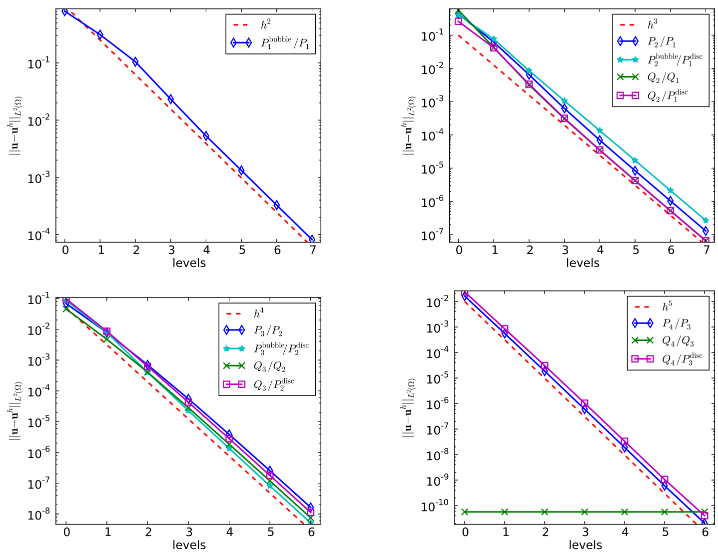

Fig. 5.3 Example 5.28. Convergence of the errors $\left\|\boldsymbol{u} - \boldsymbol{u}^h\right\|_{L^2(\Omega)}$ for different discretizations with different orders k.

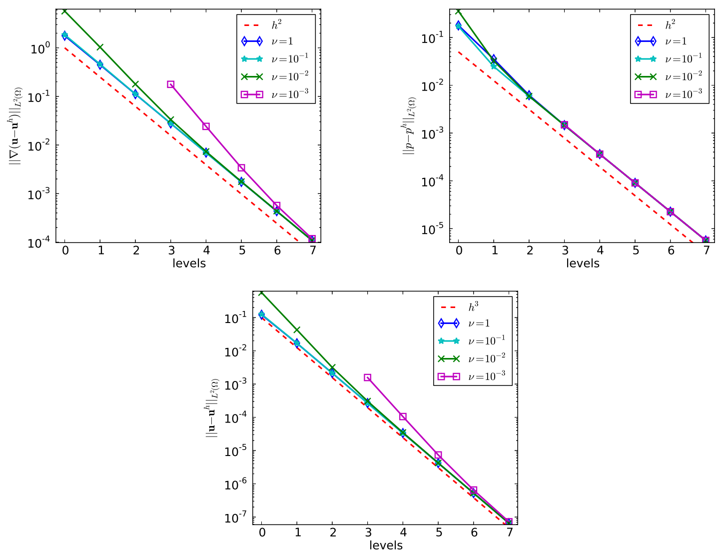

Fig. 5.4 Example 5.28. Convergence of the errors for different values of ν, Q_2/Q_1 finite element.

element discretizations. For these elements, the same functions are used for the definition of the finite element spaces and the Konstruktion of the map from the reference cell to the physical mesh cell. In this way, one obtains a better approximation of the boundary Γ_{cyl}, but not yet the correct representation. Isoparametric finite elements were used in the studies presented here. The approximation of the boundary of the cylinder is denoted by Γ_{cyl}^h. The solution of the nonlinear systems was stopped if the Euclidean norm of the residual vector was less than 10^{-15}.

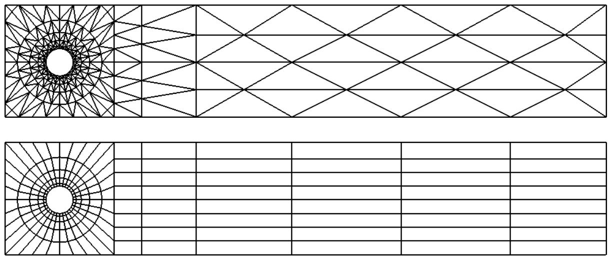

Fig. 5.5 Example 5.29. Initial grids (level 0) used for the simulations.

For computing the drag and the lift coefficient, the volume formulations (D.19) and (D.20) with the convective form of the convection term were used. In these formulations, one has to specify the functions $\boldsymbol{w}_d$ and $\boldsymbol{w}_l$ in the interior of Ω. Because these functions are up to the boundary arbitrary functions, one can use in actual computations finite element functions with appropriate boundary values. For the results presented below, the functions were chosen in such a way that they have the same order as the finite element velocity, the degrees of freedom at Γ_{cyl}^h were set to be one in the needed component, and all other degrees of freedom were set to be zero, see Figure 5.6. With this approach, only the evaluation of volume integrals in one layer of mesh cells around the circle is necessary for computing the coefficients.

The used grids possess nodes in the points $(0.15, 0.2)$ and $(0.25, 0.2)$. Thus, the finite element pressure for discretizations with discontinuous pressure approximation is in particular not continuous in these points. For computing the difference of the pressure (D.15), the values of the finite element pressure coming from all mesh cells with the node $(0.15, 0.2)$ or $(0.25, 0.2)$, respectively, were averaged.

Results for the drag coefficient, the lift coefficient, and the difference of the pressure between the front and the top of the cylinder are presented in Figures 5.7 – 5.10. It can be seen that many results show a certain order of convergence. To our best knowledge, a numerical analysis of this phenomenon is not available. A possible approach was presented in John et al. (1998). It is however not clear if the regularity assumptions on the solution of the continuous problem assumed in this paper are always true.

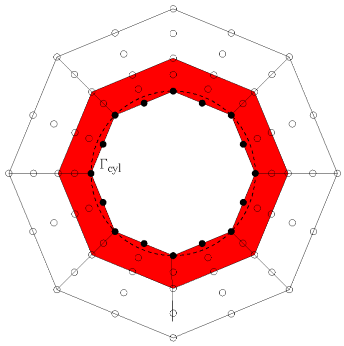

Fig. 5.6 Example 5.29. Choice of $\boldsymbol{w}_d$ and $\boldsymbol{w}_l$ for Q_2. In the filled bullet, the value 1 was set in the respective component, in the empty bullets the finite element functions have value $\mathbf{0}$. The filled mesh cells form the domain for integration in (D.19) and (D.20).

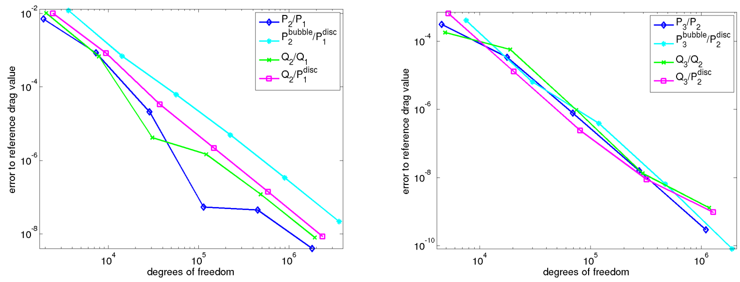

Fig. 5.7 Example 5.29. Convergence of the drag coefficient for different pairs of finite element spaces, boundary conditions (D.5), reference value (D.21). The results for boundary condition (D.4) are almost identical.

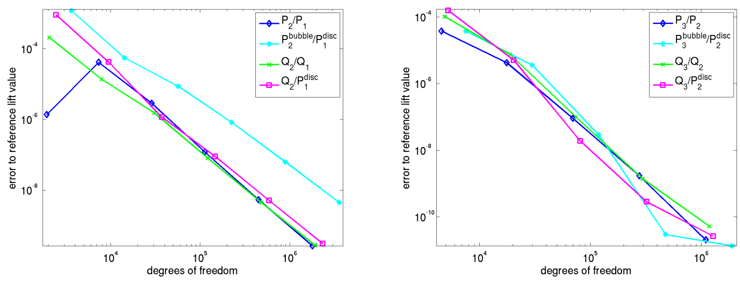

Fig. 5.8 Example 5.29. Convergence of the lift coefficient for different pairs of finite element spaces, Dirichlet boundary conditions (D.4), reference value (D.22).

Comparing the results of discretizations with different order, the higher accuracy of third order velocity/second order pressure compared with second order velocity/first order pressure can be observed also for quantities of interest which are not Sobolev norms of the error. The comparable inaccurate

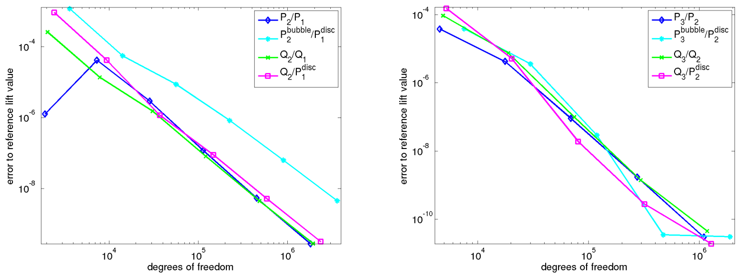

Fig. 5.9 Example 5.29. Convergence of the lift coefficient for different pairs of finite element spaces, do-nothing boundary conditions (D.5), reference value (D.24).

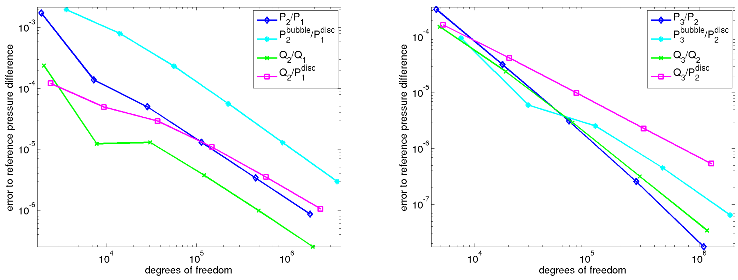

Fig. 5.10 Example 5.29. Convergence of the pressure difference for different pairs of finite element spaces, boundary conditions (D.5), reference value (D.23). The results for boundary condition (D.4) are almost identical.

results for the $P_2^{\text{bubble}}/P_1^{\text{disc}}$ discretization are notable. It can be also seen that with discontinuous pressure approximations and averaging of pressure values, comparable inaccurate results for the pressure difference are obtained, see Figure 5.10. □

Remark 5.30. Convective forms in simulations. Despite the lack of a finite element error analysis, the convective form of the convective term is used often in simulations. Sometimes, simulations with the skew-symmetric form can be found in the literature. The rotational form became somewhat popular in recent years. To our best knowledge, the divergence form is practically not used.

Comprehensive studies on the advantages and drawbacks of the different forms of the convective term were performed in Rockel (2013). todo details □

Remark 5.31. Other discretizations. nonconforming, other n, stabilizations □

5.3 Iteration Schemes for Solving the Nonlinear Problem

Remark 5.32. General fixed point iteration. The Navier–Stokes equations (5.3) can be written in operator form, see also Remark 2.4 for this concept,

$$\mathbf{0} = \boldsymbol{f} - A\boldsymbol{u} - N_{\boldsymbol{u}}\boldsymbol{u} - B'p + B\boldsymbol{u} \quad \text{in } V' \times Q',$$

where $N_{\boldsymbol{u}} : V \to V'$ is the operator for the nonlinear convective term. Applying now an injective linear operator $N_{\text{lin}}^{-1} : V' \times Q' \to V \times Q$ yields

$$\mathbf{0} = N_{\text{lin}}^{-1}\mathbf{0} = \boldsymbol{N}_{\text{lin}}^{-1}\left(\boldsymbol{f} - A\boldsymbol{u} - N_{\boldsymbol{u}}\boldsymbol{u} - B'p + B\boldsymbol{u}\right) \quad \text{in } V \times Q.$$

Then, the operator $\boldsymbol{N}_{\text{lin}}^{-1}\left(\boldsymbol{f} - A\boldsymbol{u} - N_{\boldsymbol{u}}\boldsymbol{u} - B'p + B\boldsymbol{u}\right)$ is a map from $V \times Q \to V \times Q$ and to this map, the standard approach for a fixed point iteration can be applied: Given $\left(\boldsymbol{u}^{(m)}, p^{(m)}\right) \in V \times Q$, compute $\left(\boldsymbol{u}^{(m+1)}, p^{(m+1)}\right) \in V \times Q$ by

$$\begin{pmatrix} \boldsymbol{u}^{(m+1)} \\ p^{(m+1)} \end{pmatrix} = \begin{pmatrix} \boldsymbol{u}^{(m)} \\ p^{(m)} \end{pmatrix} - \vartheta \boldsymbol{N}_{\text{lin}}^{-1} \begin{pmatrix} \langle \boldsymbol{f}, \boldsymbol{v} \rangle_{V',V} - \boldsymbol{N}\left(\boldsymbol{u}^{(m)}; \boldsymbol{u}^{(m)}, p^{(m)}\right) \\ 0 \end{pmatrix}, \tag{5.47}$$

where

$$\boldsymbol{N}\left(\boldsymbol{w}; \boldsymbol{u}, p\right) = \begin{pmatrix} a(\boldsymbol{u}, \boldsymbol{v}) + n(\boldsymbol{w}, \boldsymbol{u}, \boldsymbol{v}) + b(\boldsymbol{v}, p) \\ b(\boldsymbol{u}, q) \end{pmatrix}$$

and $\vartheta \in (0, 1]$ is a damping factor. For convenience of notation, the operators are replaced by the bilinear forms.

An iteration of form (5.47) requires the solution of a linear problem. Let $\boldsymbol{N}_{\text{lin}} : \text{range}\left(\boldsymbol{N}_{\text{lin}}^{-1}\right) \to V' \times Q'$ be the inverse operator of $\boldsymbol{N}_{\text{lin}}^{-1}$, then the linear problem has the form

$$\boldsymbol{N}_{\text{lin}} \begin{pmatrix} \delta\boldsymbol{u}^{(m+1)} \\ \delta p^{(m+1)} \end{pmatrix} = \begin{pmatrix} \langle \boldsymbol{f}, \boldsymbol{v} \rangle_{V',V} - \boldsymbol{N}\left(\boldsymbol{u}^{(m)}; \boldsymbol{u}^{(m)}, p^{(m)}\right) \\ 0 \end{pmatrix}.$$

Writing the update in the form

$$\begin{pmatrix} \delta\boldsymbol{u}^{(m+1)} \\ \delta p^{(m+1)} \end{pmatrix} = \begin{pmatrix} \tilde{\boldsymbol{u}}^{(m+1)} - \boldsymbol{u}^{(m)} \\ \tilde{p}^{(m+1)} - p^{(m)} \end{pmatrix},$$

the linear system can be reformulated for a new velocity and pressure solution

$$\boldsymbol{N}_{\text{lin}} \begin{pmatrix} \tilde{\boldsymbol{u}}^{(m+1)} \\ \tilde{p}^{(m+1)} \end{pmatrix} = \begin{pmatrix} \langle \boldsymbol{f}, \boldsymbol{v} \rangle_{V',V} - \boldsymbol{N}\left(\boldsymbol{u}^{(m)}; \boldsymbol{u}^{(m)}, p^{(m)}\right) \\ 0 \end{pmatrix} + \boldsymbol{N}_{\text{lin}} \begin{pmatrix} \boldsymbol{u}^{(m)} \\ p^{(m)} \end{pmatrix}. \tag{5.48}$$

□

Remark 5.33. Fixed point iteration with scaled Stokes equations. A simple iterative approach is obtained by setting

$$\boldsymbol{N}_{\text{lin}} = \boldsymbol{N}\left(\boldsymbol{0}; \tilde{\boldsymbol{u}}^{(m+1)}, \tilde{p}^{(m+1)}\right).$$

Inserting this expression into (5.48) gives

$$\begin{aligned}
&\begin{pmatrix} a(\tilde{\boldsymbol{u}}^{(m+1)}, \boldsymbol{v}) + b(\boldsymbol{v}, \tilde{p}^{(m+1)}) \\ b(\tilde{\boldsymbol{u}}^{(m+1)}, q) \end{pmatrix} \\
&= \begin{pmatrix} \langle \boldsymbol{f}, \boldsymbol{v}\rangle_{V',V} - a(\boldsymbol{u}^{(m)}, \boldsymbol{v}) - n(\boldsymbol{u}^{(m)}, \boldsymbol{u}^{(m)}, \boldsymbol{v}) - b(\boldsymbol{v}, p^{(m)}) \\ -b(\boldsymbol{u}^{(m)}, q) \end{pmatrix} \\
&\quad + \begin{pmatrix} a(\boldsymbol{u}^{(m)}, \boldsymbol{v}) + b(\boldsymbol{v}, p^{(m)}) \\ b(\boldsymbol{u}^{(m)}, q) \end{pmatrix} \\
&= \begin{pmatrix} \langle \boldsymbol{f}, \boldsymbol{v}\rangle_{V',V} - n(\boldsymbol{u}^{(m)}, \boldsymbol{u}^{(m)}, \boldsymbol{v}) \\ 0 \end{pmatrix}.
\end{aligned}$$

These equations are scaled Stokes equation, see (3.83).

This approach requires only the solution of a scaled Stokes problem with the same matrix and with a different right-hand side in each iteration step. It is well known for poor convergence properties in the case that ν is not sufficiently large. That means, it converges only with the application of strong damping or there is even no convergence at all if the initial iterate is not sufficiently close to the solution. Therefore, this type of fixed point iteration is in general not recommended and it will not be considered further here. □

Remark 5.34. Picard iteration. The so-called Picard iteration is obtained by setting

$$\boldsymbol{N}_{\text{lin}} = \boldsymbol{N}\left(\boldsymbol{u}^{(m)}; \tilde{\boldsymbol{u}}^{(m+1)}, \tilde{p}^{(m+1)}\right).$$

One obtains for the linear system (5.48) to be solved

$$\begin{aligned}
&\begin{pmatrix} a(\tilde{\boldsymbol{u}}^{(m+1)}, \boldsymbol{v}) + n(\boldsymbol{u}^{(m)}, \tilde{\boldsymbol{u}}^{(m+1)}, \boldsymbol{v}) + b(\boldsymbol{v}, \tilde{p}^{(m+1)}) \\ b(\tilde{\boldsymbol{u}}^{(m+1)}, q) \end{pmatrix} \\
&= \begin{pmatrix} \langle \boldsymbol{f}, \boldsymbol{v}\rangle_{V',V} - a(\boldsymbol{u}^{(m)}, \boldsymbol{v}) - n(\boldsymbol{u}^{(m)}, \boldsymbol{u}^{(m)}, \boldsymbol{v}) - b(\boldsymbol{v}, p^{(m)}) \\ -b(\boldsymbol{u}^{(m)}, q) \end{pmatrix} \\
&\quad + \begin{pmatrix} a(\boldsymbol{u}^{(m)}, \boldsymbol{v}) + n(\boldsymbol{u}^{(m)}, \boldsymbol{u}^{(m)}, \boldsymbol{v}) + b(\boldsymbol{v}, p^{(m)}) \\ b(\boldsymbol{u}^{(m)}, q) \end{pmatrix} \\
&= \begin{pmatrix} \langle \boldsymbol{f}, \boldsymbol{v}\rangle_{V',V} \\ 0 \end{pmatrix}.
\end{aligned} \tag{5.49}$$

The different forms of the convective terms on the left-hand side of (5.49) look like follows:

$$
\begin{aligned}
n_{\text{conv}}\left(\boldsymbol{u}^{(m)},\tilde{\boldsymbol{u}}^{(m+1)},\boldsymbol{v}\right) &= \left(\left(\boldsymbol{u}^{(m)}\cdot\nabla\right)\tilde{\boldsymbol{u}}^{(m+1)},\boldsymbol{v}\right),\\
n_{\text{div}}\left(\boldsymbol{u}^{(m)},\tilde{\boldsymbol{u}}^{(m+1)},\boldsymbol{v}\right) &= \left(\left(\boldsymbol{u}^{(m)}\cdot\nabla\right)\tilde{\boldsymbol{u}}^{(m+1)},\boldsymbol{v}\right)\\
&\quad+\frac{1}{2}\left(\left(\nabla\cdot\boldsymbol{u}^{(m)}\right)\tilde{\boldsymbol{u}}^{(m+1)},\boldsymbol{v}\right),\\
n_{\text{rot}}\left(\boldsymbol{u}^{(m)},\tilde{\boldsymbol{u}}^{(m+1)},\boldsymbol{v}\right) &= \left(\left(\nabla\times\boldsymbol{u}^{(m)}\right)\times\tilde{\boldsymbol{u}}^{(m+1)},\boldsymbol{v}\right),\\
n_{\text{skew}}\left(\boldsymbol{u}^{(m)},\tilde{\boldsymbol{u}}^{(m+1)},\boldsymbol{v}\right) &= \frac{1}{2}\Big[\left(\left(\boldsymbol{u}^{(m)}\cdot\nabla\right)\tilde{\boldsymbol{u}}^{(m+1)},\boldsymbol{v}\right)\\
&\quad-\left(\left(\boldsymbol{u}^{(m)}\cdot\nabla\right)\boldsymbol{v},\tilde{\boldsymbol{u}}^{(m+1)}\right)\Big].
\end{aligned}
$$

Thus, using $n_{\text{conv}}\left(\boldsymbol{u}^{(m)},\tilde{\boldsymbol{u}}^{(m+1)},\boldsymbol{v}\right)$ one has to solve in (5.49) Oseen equations with $\boldsymbol{b}=\boldsymbol{u}^{(m)}$ and $c=0$. For $n_{\text{div}}\left(\boldsymbol{u}^{(m)},\tilde{\boldsymbol{u}}^{(m+1)},\boldsymbol{v}\right)$, one obtains Oseen equations with $\boldsymbol{b}=\boldsymbol{u}^{(m)}$ and $c=\nabla\cdot\boldsymbol{u}^{(m)}$. In both cases, the equations are dominated by convection if ν is small compared with $\left\|\boldsymbol{u}^{(m)}\right\|_{L^\infty(\Omega)}$. The use of $n_{\text{rot}}\left(\boldsymbol{u}^{(m)},\tilde{\boldsymbol{u}}^{(m+1)},\boldsymbol{v}\right)$ leads to Oseen equations with $\boldsymbol{b}=\boldsymbol{0}$ and the matrix $\left(\nabla\times\boldsymbol{u}^{(m)}\right)$ as coefficient of the reactive term. In this case, there is no convection term in the linear problem.

Note that for finite element functions in all cases the assumptions on the coefficients of the Oseen equations which were made in the analysis of the Oseen equations, see Remark 4.2, are generally not fulfilled by the coefficients coming from the fixed point iteration of the stationary Navier–Stokes equations. That means, $\boldsymbol{u}^{h,(m)}$ is generally not weakly divergence-free, $\nabla\cdot\boldsymbol{u}^{h,(m)}$ might be negative, and numerical analysis for a matrix coefficients $\left(\nabla\times\boldsymbol{u}^{h,(m)}\right)$ is even not available in the literature. However, even if such an analysis can be performed, the natural extension of the assumptions on the scalar reaction coefficient to a matrix coefficient would be that the matrix is symmetric and positive semi-definite. These assumptions are generally not fulfilled by $\left(\nabla\times\boldsymbol{u}^{h,(m)}\right)$.

The skew-symmetric form of the convective term does not lead to an equation of Oseen type, since in the zeroth order term with respect to $\tilde{\boldsymbol{u}}^{(m+1)}$, the gradient of the test function appears instead of the test function itself.

Since the matrix in (5.49) depends on the current approximation $\boldsymbol{u}^{(m)}$, it changes in every iteration. □

Remark 5.35. Picard method: structure of matrices and memory requirements. After having applied an inf-sup stable finite element method, the linear system (5.49) has the saddle point form

$$
\begin{pmatrix} A & B^T \\ B & 0 \end{pmatrix}\begin{pmatrix} \underline{u} \\ \underline{p} \end{pmatrix} = \begin{pmatrix} \underline{f} \\ \underline{0} \end{pmatrix}. \tag{5.50}
$$

- *Convective form of the convective term.* In the finite element equation (5.49) that arises in the fixed point iteration for solving the non-linearity, the term

$$n_{\text{conv}}\left(\boldsymbol{u}^{h,(m)}, \tilde{\boldsymbol{u}}^{h,(m+1)}, \boldsymbol{v}^h\right) = \left(\left(\boldsymbol{u}^{h,(m)} \cdot \nabla\right) \tilde{\boldsymbol{u}}^{h,(m+1)}, \boldsymbol{v}^h\right) \qquad (5.51)$$

appears on the left-hand side. Here, $\boldsymbol{u}^{h,(m)}$ is a known finite element function. Using the ansatz (3.72) for $\tilde{\boldsymbol{u}}^{h,(m+1)}$ and considering a test function $\boldsymbol{\phi}_i^h$, one gets for the convective form of the convective term

$$\left(\left(\boldsymbol{u}^{h,(m)} \cdot \nabla\right) \tilde{\boldsymbol{u}}^{h,(m+1)}, \boldsymbol{\phi}_i^h\right) = \sum_{j=1}^{3N_v} u_j^h \left(\left(\boldsymbol{u}^{h,(m)} \cdot \nabla\right) \boldsymbol{\phi}_j^h, \boldsymbol{\phi}_i^h\right).$$

Thus, the (i,j)-matrix entry is

$$\begin{aligned} (A)_{ij} &= \int_\Omega \left(\boldsymbol{u}^{h,(m)} \cdot \nabla\right) \boldsymbol{\phi}_j^h \cdot \boldsymbol{\phi}_i^h \, d\boldsymbol{x} \\ &= \sum_{k=1}^{3} \int_\Omega \left(\left(\boldsymbol{u}^{h,(m)} \cdot \nabla\right) \boldsymbol{\phi}_j^h\right)_k \left(\boldsymbol{\phi}_i^h\right)_k \, d\boldsymbol{x}. \end{aligned} \qquad (5.52)$$

The product $\left(\left(\boldsymbol{u}^{h,(m)} \cdot \nabla\right) \boldsymbol{\phi}_j^h\right)_k \left(\boldsymbol{\phi}_i^h\right)_k$ vanishes if the k-th component of $\boldsymbol{\phi}_j^h$ or the k-th component of $\boldsymbol{\phi}_i^h$ vanishes. If $\boldsymbol{\phi}_i^h$ and $\boldsymbol{\phi}_j^h$ posses the same non-vanishing component, the matrix entry is independent of the component. Thus, the concvective term in (5.51) leads to a matrix block of the velocity-velocity couplings of the form

$$A = \begin{pmatrix} A_{11} & 0 & 0 \\ 0 & A_{11} & 0 \\ 0 & 0 & A_{11} \end{pmatrix}. \qquad (5.53)$$

- *Divergence form of the convective term.* In addition to matrix entries of form (5.52), the divergence form has the following entries to $(A)_{ij}$

$$\frac{1}{2}\int_\Omega \left(\nabla \cdot \boldsymbol{u}^{h,(m)}\right) \boldsymbol{\phi}_j^h \cdot \boldsymbol{\phi}_i^h \, d\boldsymbol{x} = \frac{1}{2}\sum_{k=1}^{3} \int_\Omega \left(\nabla \cdot \boldsymbol{u}^{h,(m)}\right) \left(\boldsymbol{\phi}_j^h\right)_k \left(\boldsymbol{\phi}_i^h\right)_k \, d\boldsymbol{x}.$$

A contribution of this type occurs also for the Galerkin discretization of the Oseen equations, see Remark 4.19. It is obvious that this entry is zero if the non-vanishing components of $\boldsymbol{\phi}_i$ and $\boldsymbol{\phi}_j$ are not the same. Otherwise, the value of this entry is independent of the index k of the non-vanishing entry. Altogether, matrix of the velocity-velocity coupling has the form (5.53).

- *Rotational form of the convective term.* The matrix entry for the rotational form is given by

$$
\begin{aligned}
&(A)_{ij} \\
&= \left(\left(\nabla \times \boldsymbol{u}^{h,(m)}\right) \times \boldsymbol{\phi}_j^h, \boldsymbol{\phi}_i^h\right) \\
&= \int_\Omega \begin{pmatrix} \left(\nabla \times \boldsymbol{u}^{h,(m)}\right)_2 \left(\boldsymbol{\phi}_j^h\right)_3 - \left(\nabla \times \boldsymbol{u}^{h,(m)}\right)_3 \left(\boldsymbol{\phi}_j^h\right)_2 \\ \left(\nabla \times \boldsymbol{u}^{h,(m)}\right)_3 \left(\boldsymbol{\phi}_j^h\right)_1 - \left(\nabla \times \boldsymbol{u}^{h,(m)}\right)_1 \left(\boldsymbol{\phi}_j^h\right)_3 \\ \left(\nabla \times \boldsymbol{u}^{h,(m)}\right)_1 \left(\boldsymbol{\phi}_j^h\right)_2 - \left(\nabla \times \boldsymbol{u}^{h,(m)}\right)_2 \left(\boldsymbol{\phi}_j^h\right)_1 \end{pmatrix} \cdot \begin{pmatrix} \left(\boldsymbol{\phi}_i^h\right)_1 \\ \left(\boldsymbol{\phi}_i^h\right)_2 \\ \left(\boldsymbol{\phi}_i^h\right)_3 \end{pmatrix} d\boldsymbol{x}.
\end{aligned}
$$

  It follows that even if the non-vanishing components of $\boldsymbol{\phi}_i^h$ and $\boldsymbol{\phi}_j^h$ are different, the resulting entry will not vanish. Hence, the matrix for the velocity-velocity coupling has the form

$$
A = \begin{pmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{pmatrix}, \tag{5.54}
$$

  which is the general form for this matrix.
- *Skew-symmetric form of the convective term.* Besides the half of the term (5.52), the skew-symmetric form possesses the contribution

$$
\begin{aligned}
&-\frac{1}{2} \int_\Omega \left(\boldsymbol{u}^{h,(m)} \cdot \nabla\right) \boldsymbol{\phi}_i^h \cdot \boldsymbol{\phi}_j^h \, d\boldsymbol{x} \\
&= -\frac{1}{2} \sum_{k=1}^{3} \int_\Omega \left(\left(\boldsymbol{u}^{h,(m)} \cdot \nabla\right) \boldsymbol{\phi}_i^h\right)_k \left(\boldsymbol{\phi}_j^h\right)_k \, d\boldsymbol{x}
\end{aligned}
$$

  in $(A)_{ij}$. From the same discussion as for (5.52), it follows that the matrix of the velocity-velocity coupling has the block-diagonal form (5.53).

Summary. Using the convective form, the divergence form, and the skew-symmetric form of the convective term leads to block-diagonal matrices of form (5.53) in the Picard iteration. Only the rotational form requires the use of a full matrix of form (5.54). □

Remark 5.36. Newton's method. In Newton's method, one takes as linear operator the derivative of the nonlinear operator at the current iterate

$$
\boldsymbol{N}_{\text{lin}} = D\boldsymbol{N} \begin{pmatrix} \boldsymbol{u}^{(m)} \\ p^{(m)} \end{pmatrix}.
$$

Considering the Gâteaux derivative, one obtains, using the linearity of $\boldsymbol{N}$ in each argument,

$$
\begin{aligned}
D\boldsymbol{N}\begin{pmatrix}\boldsymbol{u}\\ p\end{pmatrix} &= \lim_{\varepsilon\to 0}\frac{\boldsymbol{N}(\boldsymbol{u}+\varepsilon\boldsymbol{\phi};\boldsymbol{u}+\varepsilon\boldsymbol{\phi},p+\varepsilon\psi)-\boldsymbol{N}(\boldsymbol{u};\boldsymbol{u},p)}{\varepsilon}\\
&= \boldsymbol{N}(\boldsymbol{\phi};\boldsymbol{u},p)+\boldsymbol{N}(\boldsymbol{u};\boldsymbol{\phi},p)+\boldsymbol{N}(\boldsymbol{u},\boldsymbol{u},\psi).
\end{aligned}
$$

Using this operator as $\boldsymbol{N}_{\text{lin}}$ in (5.48) leads to

$$
\begin{aligned}
&\boldsymbol{N}(\tilde{\boldsymbol{u}}^{(m+1)};\boldsymbol{u}^{(m)},p^{(m)})+\boldsymbol{N}(\boldsymbol{u}^{(m)};\tilde{\boldsymbol{u}}^{(m+1)},p^{(m)})+\boldsymbol{N}(\boldsymbol{u}^{(m)},\boldsymbol{u}^{(m)},\tilde{p}^{(m+1)})\\
&\quad=\begin{pmatrix}\langle\boldsymbol{f},\boldsymbol{v}\rangle_{V',V}\\ 0\end{pmatrix}-\boldsymbol{N}(\boldsymbol{u}^{(m)};\boldsymbol{u}^{(m)},p^{(m)})+\boldsymbol{N}(\boldsymbol{u}^{(m)};\boldsymbol{u}^{(m)},p^{(m)})\\
&\qquad+\boldsymbol{N}(\boldsymbol{u}^{(m)};\boldsymbol{u}^{(m)},p^{(m)})+\boldsymbol{N}(\boldsymbol{u}^{(m)},\boldsymbol{u}^{(m)},p^{(m)}).
\end{aligned}
$$

Collecting terms gives

$$
\begin{aligned}
&\begin{pmatrix}a(\tilde{\boldsymbol{u}}^{(m+1)},\boldsymbol{v})+n(\boldsymbol{u}^{(m)},\tilde{\boldsymbol{u}}^{(m+1)},\boldsymbol{v})+n(\tilde{\boldsymbol{u}}^{(m+1)},\boldsymbol{u}^{(m)},\boldsymbol{v})+b(\boldsymbol{v},\tilde{p}^{(m+1)})\\ b(\tilde{\boldsymbol{u}}^{(m+1)},q)\end{pmatrix}\\
&\qquad=\begin{pmatrix}\langle\boldsymbol{f},\boldsymbol{v}\rangle_{V',V}+n(\boldsymbol{u}^{(m)},\boldsymbol{u}^{(m)},\boldsymbol{v})\\ 0\end{pmatrix}.
\end{aligned}
\tag{5.55}
$$

In this method, the matrix and the right-hand side change in every iteration. Problem (5.55) is an Oseen problem with $\boldsymbol{b}=\boldsymbol{u}^{(m)}$ and the tensor-valued reaction $\nabla\boldsymbol{u}^{(m)}$. □

Remark 5.37. On Newton's method.

- The order of convergence of Newton's method is expected to be better than of the Picard iteration if
  - the solution $(\boldsymbol{u},p)$ is sufficiently smooth,
  - the linear systems (5.55) are solved sufficiently accurately.

  For two-dimensional problems, there are powerful direct sparse solvers which allow for an accurate solution of the linear systems if the number of degrees of freedom is not too large, currenly $\le 5\cdot 10^5\ -\ 5\cdot 10^6$. For large two-dimensional problems and in particular for three-dimensional problems, one has to use iterative solvers. In this case, it turns out to be rather inefficient to solve the linear systems accurately. A common criterion consists in using as stopping criterion for their solution the reduction of the Euclidean norm of the residual vector by a prescribed factor, e.g., by the factor 10.
- Newton's method involves the 'reactive' term $\left(\left(\boldsymbol{u}^{(m+1)}\cdot\nabla\right)\boldsymbol{u}^{(m)},\boldsymbol{v}\right)$ on the left-hand side. This term does not fit into the theory of the Oseen equations, since the required non-negativity of this term stated in Remark 4.2 is generally not given.

  This term may lead to difficulties in the convergence of Newton's method since the properties of the tensor-valued reaction are not clear.

□

Remark 5.38. Newton's method: structure of matrices and memory requirements. The application of an inf-sup stable finite element method to (5.55) leads to a linear saddle point problem of form (5.50).

Convective form of the convective term. The term

$$
\begin{aligned}
& n\left(\boldsymbol{u}^{h,(m)}, \boldsymbol{u}^{h,(m+1)}, \boldsymbol{v}^h\right) + n\left(\boldsymbol{u}^{h,(m+1)}, \boldsymbol{u}^{(h,m)}, \boldsymbol{v}^h\right) \\
& = \left(\left(\boldsymbol{u}^{h,(m)} \cdot \nabla\right) \boldsymbol{u}^{h,(m+1)}, \boldsymbol{v}^h\right) + \left(\left(\boldsymbol{u}^{h,(m+1)} \cdot \nabla\right) \boldsymbol{u}^{h,(m)}, \boldsymbol{v}^h\right)
\end{aligned}
$$

arises on the left-hand side of the equation. The corresponding matrix entry for the convective form of the convective term becomes

$$
\begin{aligned}
(A)_{ij} = \sum_{k=1}^{3} \int_\Omega \Bigg[& \left(\boldsymbol{u}^{h,(m)} \cdot \nabla \left(\boldsymbol{\phi}_j^h\right)\right)_k \cdot \left(\boldsymbol{\phi}_i^h\right)_k \\
& + \left(\sum_{l=1}^{3} \left(\boldsymbol{\phi}_j^h\right)_l \cdot \nabla \left(\boldsymbol{u}^{h,(m)}\right)_l\right)_k \cdot \left(\boldsymbol{\phi}_i^h\right)_k \Bigg] \, d\boldsymbol{x}, \quad i,j = 1,\ldots,dN_v.
\end{aligned}
$$

The sum in the second term is generally not zero for each index k since generally $\nabla\left(\boldsymbol{u}^{h,(m)}\right)_l$ does not vanish. It follows that the second term in general does not vanish for each pair of indices (i,j) with $\left|\text{supp}\left(\boldsymbol{\phi}_i^h\right) \cap \text{supp}\left(\boldsymbol{\phi}_j^h\right)\right| > 0$. Hence, the matrix blocks A_{kl} posses in general non-zero entries for all pairs (k,l). Altogether, the matrix which originates from the convective and reactive term in (5.55) has the block form (5.54), where the block are in general mutually different since different derivatives of $\boldsymbol{u}^{h,(m)}$ have to be considered in the assembling of each block.

Divergence form and skew-symmetric form of the convective term. Both forms contain the term from the convective form. Since this term leads already to matrix of the velocity-velocity coupling of form (5.54), also the matrices for these to forms have the form (5.54).

Rotational form of the convective term. Already for the Picard iteration, the rotational form of the convective term requires the general form (5.54) of a matrix for the velocity-velocity. With the additional term which is introduced from Newton's method, one gets the same form.

Summary. The application of Newton's method leads always to a matrix for the velocity-velocity coupling of form (5.54) with mutually different blocks. □

Remark 5.39. Memory requirements and computational costs for the complete linear saddle point problem. The matrix of the velocity-velocity couplings in the linear saddle point problems (5.49) and (5.55) is the sum of the matrix which arises in the discretization of the viscous term, see (3.76), and the matrix from the linearization of the convective term. The number of blocks in the velocity-velocity couplings that has to be stored is given in Table 5.1. It can be seen that using Newton's method requires the storage of more velocity-

velocity matrix blocks. As a consequence, the computational costs for matrix assembling are larger for Newton's method. As there are more non-vanishing blocks, more operations are necessary in performing matrix-vector products.

Table 5.1 Number of matrix blocks to be stored in solving the three-dimensional Navier–Stokes equations, convective form of the nonlinear term.

form of viscous term	fixed point (5.49)	Newton (5.55)
$\left(\nu\nabla\boldsymbol{u}^{h,(m+1)},\nabla\boldsymbol{v}^h\right)$	1	9
$\left(2\nu\mathbb{D}\left(\boldsymbol{u}^{h,(m+1)}\right),\mathbb{D}\left(\boldsymbol{v}^h\right)\right)$	6	9

□

5.4 A Posteriori Error Estimates

Remark 5.40. Goals of a posteriori error estimators. So far, so called a priori error estimates were proved, e.g., in Corollary 3.28. The right-hand side of such estimates depends on the unknown solution of the continuous problem and on an unknown constant.

The goals of a posteriori error estimates are twofold. First, they should provide estimates of errors between a computed solution $\left(\boldsymbol{u}^h, p^h\right)$ and the unknown solution $(\boldsymbol{u}, p)$ of the continuous problem (3.2). Usually, the errors are measured in norms of Sobolev spaces defined on Ω. That means, estimates of global errors from above have to be derived, which have the form, e.g., for the velocity

$$\left\|\boldsymbol{u}-\boldsymbol{u}^h\right\|_\Omega \le C\eta. \tag{5.56}$$

In (5.56), η is a quantity which is computable with the information available in the numerical solution process and C is a positive constant which should be independent of the mesh width and the solution and for which one should have an idea of its order of magnitude.

The second task of a posteriori error estimates consists in controlling an adaptive mesh refinement. The rationale behind this idea is that the error of the computed solution to the solution of the continuous problem can be reduced best, or at least significantly, if the mesh in those subregions of the domain is refined, where the local error is largest. Then, a local a posteriori error estimate should identify these subregions. To this end, a local lower estimate from below of the form

$$\eta_K \le C\left\|\boldsymbol{u}-\boldsymbol{u}^h\right\|_{\omega(K)} \tag{5.57}$$

is necessary, where $\omega(K)$ denotes a small neighborhood of a mesh cell K and η_K is a computable quantity. For (5.57), it has to be proved that the positive constant C can be bounded from below and above independently of K. Estimate (5.57) tells that in subregions where the local error estimate η_K is large, also the local error is large.

There are different ways for computing a posteriori error estimates. Among the most popular ones are residual-based estimates, which where presented for the Stokes equations the first time in Verfürth (1989), and estimates which are based on the solution of local problems, that were also introduced in Verfürth (1989).

A review of residual-based estimators can be found in Verfürth (1995, 2013). □

5.4.1 A Residual-Based A Posteriori Error Estimator

Remark 5.41. Basic assumptions. This section considers conforming inf-sup stable pairs of finite element spaces, which are defined on a quasi-uniform family of triangulations. □

Lemma 5.42. Estimate of the supremum of the bilinear form. *For all* $(\boldsymbol{w}, r) \in V \times Q$ *it is*

$$\begin{aligned} &\|\nabla \boldsymbol{w}\|_{L^2(\Omega)} + \|r\|_{L^2(\Omega)} \\ &\le \frac{1}{\beta_{\mathrm{is,Bab}}} \sup_{\substack{(\boldsymbol{v},q)\in V\times Q\\ (\boldsymbol{v},q)\neq(\boldsymbol{0},0)}} \frac{(\nabla \boldsymbol{w}, \nabla \boldsymbol{v}) - (\nabla\cdot\boldsymbol{v}, r) + (\nabla\cdot\boldsymbol{w}, q)}{\|\nabla \boldsymbol{v}\|_{L^2(\Omega)} + \|q\|_{L^2(\Omega)}} \\ &\le \frac{1}{\beta_{\mathrm{is,Bab}}} \left(\|\nabla \boldsymbol{w}\|_{L^2(\Omega)} + \|r\|_{L^2(\Omega)}\right). \end{aligned} \tag{5.58}$$

Proof. The first estimate follows directly from the Babuška inf-sup condition (3.7). For the second estimate, one applies the Cauchy–Schwarz inequality (A.8) and (2.39), which gives

$$\begin{aligned} &\sup_{\substack{(\boldsymbol{v},q)\in V\times Q\\ (\boldsymbol{v},q)\neq(\boldsymbol{0},0)}} \frac{(\nabla \boldsymbol{w}, \nabla \boldsymbol{v}) - (\nabla\cdot\boldsymbol{v}, r) + (\nabla\cdot\boldsymbol{w}, q)}{\|\nabla \boldsymbol{v}\|_{L^2(\Omega)} + \|q\|_{L^2(\Omega)}} \\ &\le \sup_{\substack{(\boldsymbol{v},q)\in V\times Q\\ (\boldsymbol{v},q)\neq(\boldsymbol{0},0)}} \frac{\|\nabla \boldsymbol{w}\|_{L^2(\Omega)} \|\nabla \boldsymbol{v}\|_{L^2(\Omega)} + \|\nabla \boldsymbol{v}\|_{L^2(\Omega)} \|r\|_{L^2(\Omega)} + \|\nabla \boldsymbol{w}\|_{L^2(\Omega)} \|q\|_{L^2(\Omega)}}{\|\nabla \boldsymbol{v}\|_{L^2(\Omega)} + \|q\|_{L^2(\Omega)}} \\ &\le \sup_{\substack{(\boldsymbol{v},q)\in V\times Q\\ (\boldsymbol{v},q)\neq(\boldsymbol{0},0)}} \frac{\left(\|\nabla \boldsymbol{w}\|_{L^2(\Omega)} + \|r\|_{L^2(\Omega)}\right)\left(\|\nabla \boldsymbol{v}\|_{L^2(\Omega)} + \|q\|_{L^2(\Omega)}\right)}{\|\nabla \boldsymbol{v}\|_{L^2(\Omega)} + \|q\|_{L^2(\Omega)}} \\ &= \|\nabla \boldsymbol{w}\|_{L^2(\Omega)} + \|r\|_{L^2(\Omega)}. \end{aligned}$$

■

Remark 5.43. To estimate (5.58). Estimate (5.58) states the fact that the operator in the numerator defines an isomorphism from $V \times Q$ to $V' \times Q'$. The left estimate gives injectivity of the operator since if two different arguments would give the same results, the term in the middle of (5.58) vanishes for the difference and the left-hand side does not, which is a contradiction. The right estimate holds for all arguments and thus it just gives the boundedness of the operator. Altogether, an estimate of type (5.58) is not special for the Stokes equations but an estimate of this type holds always for well-posed problems. □

Corollary 5.44. Error estimate with the residual. *Let $V^h \subset V$ and $Q^h \subset Q$, let $(\boldsymbol{u}, p)$ be the solution of* (3.2) *and $\left(\boldsymbol{u}^h, p^h\right) \in V^h \times Q^h$ be arbitrary, then it holds*

$$
\begin{aligned}
&\sup_{\substack{(\boldsymbol{v},q)\in V\times Q\\(\boldsymbol{v},q)\neq(\boldsymbol{0},0)}} \frac{(\boldsymbol{f},\boldsymbol{v}) - (\nabla \boldsymbol{u}^h, \nabla \boldsymbol{v}) + (\nabla\cdot\boldsymbol{v}, p^h) - (\nabla\cdot\boldsymbol{u}^h, q)}{\|\nabla \boldsymbol{v}\|_{L^2(\Omega)} + \|q\|_{L^2(\Omega)}}\\
&\le \left\|\nabla(\boldsymbol{u}-\boldsymbol{u}^h)\right\|_{L^2(\Omega)} + \left\|p-p^h\right\|_{L^2(\Omega)} \qquad (5.59)\\
&\le \frac{1}{\beta_{\mathrm{is,Bab}}} \sup_{\substack{(\boldsymbol{v},q)\in V\times Q\\(\boldsymbol{v},q)\neq(\boldsymbol{0},0)}} \frac{(\boldsymbol{f},\boldsymbol{v}) - (\nabla \boldsymbol{u}^h, \nabla \boldsymbol{v}) + (\nabla\cdot\boldsymbol{v}, p^h) - (\nabla\cdot\boldsymbol{u}^h, q)}{\|\nabla \boldsymbol{v}\|_{L^2(\Omega)} + \|q\|_{L^2(\Omega)}}.
\end{aligned}
$$

Proof. Setting $\boldsymbol{w} = \boldsymbol{u} - \boldsymbol{u}^h$ and $r = p - p^h$ in (5.58) gives for the numerator

$$
\begin{aligned}
&(\nabla(\boldsymbol{u}-\boldsymbol{u}^h), \nabla \boldsymbol{v}) - (\nabla\cdot\boldsymbol{v}, p-p^h) + (\nabla\cdot(\boldsymbol{u}-\boldsymbol{u}^h), q)\\
&= (\boldsymbol{f},\boldsymbol{v}) - (\nabla \boldsymbol{u}^h, \nabla \boldsymbol{v}) + (\nabla\cdot\boldsymbol{v}, p^h) - (\nabla\cdot\boldsymbol{u}^h, q). \qquad (5.60)
\end{aligned}
$$

Now, the statement of the corollary follows from (5.58). ■

Remark 5.45. To estimate (5.59). The supremum in estimate (5.59) is just the definition of the norm of the residual in $V' \times Q'$. Hence estimate (5.59) states that the error $\left\|\nabla(\boldsymbol{u}-\boldsymbol{u}^h)\right\|_{L^2(\Omega)} + \left\|p-p^h\right\|_{L^2(\Omega)}$ for an arbitrary pair $\left(\boldsymbol{u}^h, p^h\right) \in V^h \times Q^h$ is bounded from below and from above by the norm of the residual in the dual space. However, in practice one cannot compute this norm since the supremum is taken in a infinite-dimensional space. The goal consists now in estimating this norm with computable expressions. □

Theorem 5.46. Global upper, residual-based, a posteriori error estimate for conforming inf-sup stable finite element spaces. *Let $\boldsymbol{f} \in L^2(\Omega)$, let $P^h\boldsymbol{f}$ a polynomial approximation of $\boldsymbol{f}$ (which can be integrated exactly), and consider conforming finite element spaces V^h/Q^h which satisfy the discrete inf-sup condition* (2.45) *on a quasi-uniform family of triangulations $\{\mathcal{T}^h\}_{h>0}$. Let $(\boldsymbol{u}, p)$ be the solution of* (3.2) *and $\left(\boldsymbol{u}^h, p^h\right)$ be the solution of* (3.13). *Defining the mesh cell residual*

$$
\boldsymbol{r}_K^h\left(\boldsymbol{u}^h, p^h\right) = \left(P^h\boldsymbol{f} + \Delta\boldsymbol{u}^h - \nabla p^h\right)\big|_K \quad \forall\, K \in \mathcal{T}^h, \qquad (5.61)
$$

the face residual

$$\boldsymbol{r}_E^h\left(\boldsymbol{u}^h,p^h\right)=\left[\left|\left(-\nabla\boldsymbol{u}^h+p^h\mathbb{I}\right)\boldsymbol{n}_E\right|\right]_E\quad\forall\,E\in\mathcal{E}^h,\tag{5.62}$$

and the local error estimator

$$\eta_K=\left(h_K^2\left\|\boldsymbol{r}_K^h\left(\boldsymbol{u}^h,p^h\right)\right\|_{L^2(K)}^2+\left\|\nabla\cdot\boldsymbol{u}^h\right\|_{L^2(K)}^2\right.$$
$$\left.+\frac{1}{2}\sum_{E\subset\partial K,E\in\mathcal{E}^h}h_E\left\|\boldsymbol{r}_E^h\left(\boldsymbol{u}^h,p^h\right)\right\|_{L^2(E)}^2\right)^{1/2},$$

then it holds the a posteriori estimate

$$\begin{aligned}&\left\|\nabla(\boldsymbol{u}-\boldsymbol{u}^h)\right\|_{L^2(\Omega)}+\left\|p-p^h\right\|_{L^2(\Omega)}\\&\le C\left(\sum_{K\in\mathcal{T}^h}\eta_K^2+h_K^2\left\|\boldsymbol{f}-P^h\boldsymbol{f}\right\|_{L^2(K)}^2\right)^{1/2},\end{aligned}\tag{5.63}$$

where the constant does not depend on the solution and on the mesh width.

Proof. Subtracting the finite element equation (3.13) from the weak form of the Stokes equations (3.2) and using $\nabla\cdot\boldsymbol{u}=0$ gives the error equation

$$\left(\nabla\left(\boldsymbol{u}-\boldsymbol{u}^h\right),\nabla\boldsymbol{v}^h\right)-\left(\nabla\cdot\boldsymbol{v}^h,p-p^h\right)+\left(\nabla\cdot\left(\boldsymbol{u}-\boldsymbol{u}^h\right),q^h\right)=0$$

for all $\left(\boldsymbol{v}^h,\boldsymbol{q}^h\right)\in V^h\times Q^h$. Using (5.60) and this equation, one gets by applying integration by parts and using $\nabla\cdot\boldsymbol{u}=0$

$$\begin{aligned}&(\boldsymbol{f},\boldsymbol{v})-(\nabla\boldsymbol{u}^h,\nabla\boldsymbol{v})+(\nabla\cdot\boldsymbol{v},p^h)-(\nabla\cdot\boldsymbol{u}^h,q)\\&=\left(\nabla\left(\boldsymbol{u}-\boldsymbol{u}^h\right),\nabla\left(\boldsymbol{v}-\boldsymbol{v}^h\right)\right)-\left(\nabla\cdot\left(\boldsymbol{v}-\boldsymbol{v}^h\right),p-p^h\right)+\left(\nabla\cdot\left(\boldsymbol{u}-\boldsymbol{u}^h\right),q-q^h\right)\\&=\left(\nabla\boldsymbol{u},\nabla\left(\boldsymbol{v}-\boldsymbol{v}^h\right)\right)-\left(\nabla\cdot\left(\boldsymbol{v}-\boldsymbol{v}^h\right),p\right)\\&\quad-\sum_{K\in\mathcal{T}^h}\int_{\partial K}\left(\nabla\boldsymbol{u}^h\boldsymbol{n}_{\partial K}\right)\cdot\left(\boldsymbol{v}-\boldsymbol{v}^h\right)\,d\boldsymbol{s}+\sum_{K\in\mathcal{T}^h}\left(\Delta\boldsymbol{u}^h,\boldsymbol{v}-\boldsymbol{v}^h\right)_K\\&\quad+\sum_{K\in\mathcal{T}^h}\int_{\partial K}p^h\left(\boldsymbol{v}-\boldsymbol{v}^h\right)\cdot\boldsymbol{n}_{\partial K}\,d\boldsymbol{s}-\sum_{K\in\mathcal{T}^h}\left(\nabla p^h,\boldsymbol{v}-\boldsymbol{v}^h\right)_K-\left(\nabla\cdot\boldsymbol{u}^h,q-q^h\right)\end{aligned}$$

for all $(\boldsymbol{v},q)\in V\times Q$ and all $\left(\boldsymbol{v}^h,\boldsymbol{q}^h\right)\in V^h\times Q^h$. Taking now $\boldsymbol{v}^h=P_{\mathrm{Cle}}^h\boldsymbol{v}$ the Clément interpolant of $\boldsymbol{v}$ which preserves homogeneous Dirichlet boundary conditions, see Remark C.22, using that $(\boldsymbol{u},p)$ solves the Stokes equations, writing the integrals on the faces with jumps, and using that $\boldsymbol{u}^h$ is discretely divergence-free leads to

$$\begin{aligned}&(\boldsymbol{f},\boldsymbol{v})-(\nabla\boldsymbol{u}^h,\nabla\boldsymbol{v})+(\nabla\cdot\boldsymbol{v},p^h)-(\nabla\cdot\boldsymbol{u}^h,q)\\&=\sum_{K\in\mathcal{T}^h}\left[\left(P^h\boldsymbol{f}+\Delta\boldsymbol{u}^h-\nabla p^h,\boldsymbol{v}-P_{\mathrm{Cle}}^h\boldsymbol{v}\right)_K+\left(\boldsymbol{f}-P^h\boldsymbol{f},\boldsymbol{v}-P_{\mathrm{Cle}}^h\boldsymbol{v}\right)_K-\left(\nabla\cdot\boldsymbol{u}^h,q\right)_K\right]\\&\quad+\sum_{E\in\mathcal{E}^h}\left(\left[\left|\left(-\nabla\boldsymbol{u}^h+p^h\mathbb{I}\right)\boldsymbol{n}_E\right|\right]_E,\boldsymbol{v}-P_{\mathrm{Cle}}^h\boldsymbol{v}\right)_E,\end{aligned}\tag{5.64}$$

where the jumps are defined in Remark 2.52. Note that for edges on the Dirichlet boundary it is $\boldsymbol{v}=P_{\mathrm{Cle}}^h\boldsymbol{v}$. In the next step, all terms are estimated with the Cauchy–Schwarz inequality (A.8) and the interpolation estimate (C.7)

$$\begin{aligned}\sum_{K\in\mathcal{T}^h} &\left(P^h\boldsymbol{f}+\Delta\boldsymbol{u}^h-\nabla p^h,\boldsymbol{v}-P^h_{\mathrm{Cle}}\boldsymbol{v}\right)_K \\ &\le C\sum_{K\in\mathcal{T}^h} h_K\left\|P^h\boldsymbol{f}+\Delta\boldsymbol{u}^h-\nabla p^h\right\|_{L^2(K)}\|\nabla\boldsymbol{v}\|_{L^2(K)} \\ &\le C\left(\sum_{K\in\mathcal{T}^h} h_K^2\left\|\boldsymbol{r}_K^h\left(\boldsymbol{u}^h,p^h\right)\right\|^2_{L^2(K)}\right)^{1/2}\|\nabla\boldsymbol{v}\|_{L^2(\Omega)},\end{aligned}$$

$$\sum_{K\in\mathcal{T}^h}\left(\boldsymbol{f}-P^h\boldsymbol{f},\boldsymbol{v}-P^h_{\mathrm{Cle}}\boldsymbol{v}\right)_K\le C\left(\sum_{K\in\mathcal{T}^h} h_K^2\left\|\boldsymbol{f}-P^h\boldsymbol{f}\right\|^2_{L^2(K)}\right)^{1/2}\|\nabla\boldsymbol{v}\|_{L^2(\Omega)},$$

$$\sum_{K\in\mathcal{T}^h}\left(\nabla\cdot\boldsymbol{u}^h,q\right)_K\le\left(\sum_{K\in\mathcal{T}^h}\left\|\nabla\cdot\boldsymbol{u}^h\right\|^2_{L^2(K)}\right)^{1/2}\|q\|_{L^2(\Omega)},$$

and with the interpolation estimate (**??**)

$$\begin{aligned}\sum_{E\in\mathcal{E}^h}&\left(\left[\left|\left(-\nabla\boldsymbol{u}^h+p^h\mathbb{I}\right)\boldsymbol{n}_E\right|\right]_E,\boldsymbol{v}-P^h_{\mathrm{Cle}}\boldsymbol{v}\right)_E \\ &\le C\left(\sum_{E\in\mathcal{E}^h} h_E\left\|\boldsymbol{r}_E^h\left(\boldsymbol{u}^h,p^h\right)\right\|^2_{L^2(E)}\right)^{1/2}\|\nabla\boldsymbol{v}\|_{L^2(\Omega)}.\end{aligned}$$

Inserting all estimates into (5.64) and observing that interior faces belong to two mesh cells leads to

$$\begin{aligned}&(\boldsymbol{f},\boldsymbol{v})-(\nabla\boldsymbol{u}^h,\nabla\boldsymbol{v})+(\nabla\cdot\boldsymbol{v},p^h)-(\nabla\cdot\boldsymbol{u}^h,q) \\ &\le C\Bigg(\sum_{K\in\mathcal{T}^h} h_K^2\left\|\boldsymbol{r}_K^h\left(\boldsymbol{u}^h,p^h\right)\right\|^2_{L^2(K)}+\left\|\nabla\cdot\boldsymbol{u}^h\right\|^2_{L^2(K)}+h_K^2\left\|\boldsymbol{f}-P^h\boldsymbol{f}\right\|^2_{L^2(K)} \\ &\quad+\frac{1}{2}\sum_{E\subset\partial K,E\in\mathcal{E}^h} h_E\left\|\boldsymbol{r}_E^h\left(\boldsymbol{u}^h,p^h\right)\right\|^2_{L^2(E)}\Bigg)^{1/2}\left(\|\nabla\boldsymbol{v}\|_{L^2(\Omega)}+\|q\|_{L^2(\Omega)}\right).\end{aligned}$$

Now, the a posteriori error estimate (5.63) follows directly from (5.59). ■

Remark 5.47. On the global upper estimate.

- The constant on the right-hand side of (5.63) depends on the constants of the local interpolation estimates (C.7), (**??**) and on $\beta_{\mathrm{is,Bab}}^{-1}$. The inf-sup constant is related to the stability of the problem. Studies of the size of constants in interpolation estimates can be found, e.g., in Carstensen and Funken (2000). more details
- For problems with non-homogeneous Dirichlet boundary conditions, an interpolation operator is used in the analysis, instead of the Clément operator, which preserves these condition, usually the Scott–Zhang operator.
- In the case of do-nothing or homogeneous natural boundary conditions (1.30) on some part Γ_{outf} of the boundary, the error estimator η_K has to be extended by the term

$$\sum_{E\subset\partial K, E\subset\Gamma_{\mathrm{outf}}} h_E \left\| \boldsymbol{r}_E^h \left(\boldsymbol{u}^h, p^h\right)\right\|_{L^2(E)}^2$$

within the parentheses, where the jump in $\boldsymbol{r}_E^h \left(\boldsymbol{u}^h, p^h\right)$ is just the difference of the boundary condition satisfied by the finite element approximation and the homogeneous boundary condition prescribed for the continuous problem.

- One can find the definition of the local estimator η_K also without the factor $1/2$ in front of the edge residuals. citation? This change of η_K changes only the constant in the estimate (5.63).

□

Remark 5.48. Principal way for obtaining a local lower estimate of form (5.57). For obtaining a local lower estimate of form (5.57), appropriate test functions are used in the continuous Stokes equations (3.2). These functions are defined with the help of mesh cell bubble functions and edge bubble functions. □

Lemma 5.49. Local estimates for bubble functions. *Let $\phi_K^h(\boldsymbol{x})$ be a cell bubble function, i.e., $\phi_K(\boldsymbol{x})$ is polynomial which is positive in the interior of K, which vanishes on ∂K and whose support is K. Then, for all polynomials $v^h(\boldsymbol{x})$ on K the following estimates hold*

$$C^{-1}\left\|v^h\right\|_{L^2(K)}^2 \le \left(v^h, v^h\phi_K^h\right)_K \le C\left\|v^h\right\|_{L^2(K)}^2, \tag{5.65}$$

$$\begin{aligned} C^{-1}\left\|v^h\right\|_{L^2(K)} &\le \left\|v^h\phi_K^h\right\|_{L^2(K)} + h_K\left\|\nabla(v^h\phi_K^h)\right\|_{L^2(K)} \\ &\le C\left\|v^h\right\|_{L^2(K)}, \end{aligned} \tag{5.66}$$

where C is independent of v^h and h_K.

Let $\phi_E^h(\boldsymbol{x})$ be a face bubble function, i.e., $\phi_E^h(\boldsymbol{x})$ is continuous, it is a polynomial on both mesh cells K_1, K_2 which share the face E, it is positive in $\omega_E = K_1 \cup K_2$, it vanishes on the boundary of ω_E, and its support is ω_E. Then, for all polynomials v^h on K_i, $i \in \{1,2\}$, one has the estimates

$$C^{-1}\left\|v^h\right\|_{L^2(E)}^2 \le \left(v^h, v^h\phi_E^h\right)_E \le C\left\|v^h\right\|_{L^2(E)}^2, \tag{5.67}$$

$$h_{K_i}^{-1/2}\left\|v^h\phi_E^h\right\|_{L^2(K_i)} + h_{K_i}^{1/2}\left\|\nabla(v^h\phi_E^h)\right\|_{L^2(K_i)} \le C\left\|v^h\right\|_{L^2(E)}, \tag{5.68}$$

where C is independent of v^h and h_{K_i}.

Proof. The proof of these estimates can be found in (Verfürth, 1994, Theorem 2.2) and (Verfürth, 1994, Theorem 2.4). work out? ■

Lemma 5.50. Local estimate for the mesh cell residual. *Under the assumption of Theorem 5.46, it is*

$$h_K \left\| \boldsymbol{r}_K^h \left(\boldsymbol{u}^h, p^h\right)\right\|_{L^2(K)} \le C\Big(\left\|\nabla(\boldsymbol{u}-\boldsymbol{u}^h)\right\|_{L^2(K)} + \left\|p-p^h\right\|_{L^2(K)} + h_K \left\|\boldsymbol{f} - P^h\boldsymbol{f}\right\|_{L^2(K)}\Big), \tag{5.69}$$

with C independent of the mesh width and the solution.

Proof. Considering a vector-valued version of (5.65) and choosing $\boldsymbol{r}_K^h\left(\boldsymbol{u}^h, p^h\right)$ as polynomial, which will be abbreviated in the proof by $\boldsymbol{r}_K^h$, gives

$$\left\|\boldsymbol{r}_K^h\right\|_{L^2(K)}^2 \le C\left(\boldsymbol{r}_K^h, \phi_K^h \boldsymbol{r}_K^h\right)_K. \tag{5.70}$$

Now, $\phi_K^h \boldsymbol{r}_K^h$ can be extended off K by zero which gives a function in V and which will be denoted by the same symbol. Using $\left(\phi_K^h \boldsymbol{r}_K^h, 0\right)$ as test function in (3.2) gives

$$\left(\nabla \boldsymbol{u}, \nabla\left(\phi_K^h \boldsymbol{r}_K^h\right)\right) - \left(\nabla\cdot\left(\phi_K^h \boldsymbol{r}_K^h\right), p\right) = \left(\boldsymbol{f}, \phi_K^h \boldsymbol{r}_K^h\right).$$

Adding and subtracting terms and applying integration by parts, using the definition (5.61) of the mesh cell residual, and observing that $\phi_K^h \boldsymbol{r}_K^h$ vanishes on ∂K yields

$$\begin{aligned}
&\left(\nabla(\boldsymbol{u}-\boldsymbol{u}^h), \nabla(\phi_K^h \boldsymbol{r}_K^h)\right) - \left(\nabla\cdot\left(\phi_K^h \boldsymbol{r}_K^h\right), p-p^h\right)\\
&= \left(P^h\boldsymbol{f}, \phi_K^h \boldsymbol{r}_K^h\right) + \left(\boldsymbol{f}-P^h\boldsymbol{f}, \phi_K^h \boldsymbol{r}_K^h\right) - \left(\nabla \boldsymbol{u}^h, \nabla(\phi_K^h \boldsymbol{r}_K^h\right) + \left(\nabla\cdot\left(\phi_K^h \boldsymbol{r}_K^h\right), p^h\right)\\
&= (\boldsymbol{r}_K^h, \phi_K^h \boldsymbol{r}_K^h)_K + \left(\boldsymbol{f}-P^h\boldsymbol{f}, \phi_K^h \boldsymbol{r}_K^h\right)_K.
\end{aligned}$$

Inserting this identity into (5.70) applying the Cauchy–Schwarz inequality (A.8), (2.39), and (5.66) leads to

$$\begin{aligned}
&\left\|\boldsymbol{r}_K^h\right\|_{L^2(K)}^2\\
&\le C\Big(\left\|\nabla(\boldsymbol{u}-\boldsymbol{u}^h)\right\|_{L^2(K)} \left\|\nabla(\phi_K^h \boldsymbol{r}_K^h)\right\|_{L^2(K)} + \left\|p-p^h\right\|_{L^2(K)} \left\|\nabla(\phi_K^h \boldsymbol{r}_K^h)\right\|_{L^2(K)}\\
&\quad + \left\|\boldsymbol{f}-P^h\boldsymbol{f}\right\|_{L^2(K)} \left\|\phi_K^h \boldsymbol{r}_K^h\right\|_{L^2(K)}\Big)\\
&\le C\left(h_K^{-1}\left\|\nabla(\boldsymbol{u}-\boldsymbol{u}^h)\right\|_{L^2(K)} + h_K^{-1}\left\|p-p^h\right\|_{L^2(K)} + \left\|\boldsymbol{f}-P^h\boldsymbol{f}\right\|_{L^2(K)}\right)\left\|\boldsymbol{r}_K^h\right\|_{L^2(K)}.
\end{aligned}$$

Dividing by $h_K^{-1}\left\|\boldsymbol{r}_K^h\right\|_{L^2(K)}$ gives the statement of the lemma. ■

Lemma 5.51. Local estimate for the face residual. *With the assumption of Theorem 5.46, it is*

$$h_E^{1/2} \left\| \boldsymbol{r}_E^h \left(\boldsymbol{u}^h, p^h\right)\right\|_{L^2(E)} \le C\Big(\left\|\nabla(\boldsymbol{u}-\boldsymbol{u}^h)\right\|_{L^2(\omega_E)} + \left\|p-p^h\right\|_{L^2(\omega_E)} + h_E \left\|\boldsymbol{f} - P^h\boldsymbol{f}\right\|_{L^2(\omega_E)}\Big), \tag{5.71}$$

with C independent of the mesh width and the solution.

Proof. Taking the face residual (5.62), abbreviating this function with $\boldsymbol{r}_E^h$, and using a vector-valued version of (5.67) gives

$$\left\|\boldsymbol{r}_E^h\right\|_{L^2(E)}^2 \le C\left(\boldsymbol{r}_E^h, \phi_E^h \boldsymbol{r}_E^h\right)_E. \tag{5.72}$$

This function $\phi_E^h \boldsymbol{r}_E^h$ can be extended to a function in V, which is denoted with the same symbol, by setting $\phi_E^h \boldsymbol{r}_E^h$ outside ω_E to zero. Now, the test function $\left(\phi_E^h \boldsymbol{r}_E^h, 0\right)$ is applied in the Stokes equations (3.2) yielding

$$\left(\nabla \boldsymbol{u}, \nabla\left(\phi_E^h \boldsymbol{r}_E^h\right)\right)-\left(\nabla \cdot\left(\phi_E^h \boldsymbol{r}_E^h\right), p\right)=\left(\boldsymbol{f}, \phi_E^h \boldsymbol{r}_E^h\right).$$

Adding and subtracting terms and applying integration by parts, noting that $\phi_E^h \boldsymbol{r}_E^h$ vanishes on the boundary of ω_E and outside ω_E leads to

$$\begin{aligned}
&\left(\nabla(\boldsymbol{u}-\boldsymbol{u}^h), \nabla(\phi_E^h \boldsymbol{r}_E^h)\right)-\left(\nabla \cdot\left(\phi_E^h \boldsymbol{r}_E^h\right), p-p^h\right) \\
&=\left(P^h \boldsymbol{f}, \phi_E^h \boldsymbol{r}_E^h\right)+\left(\boldsymbol{f}-P^h \boldsymbol{f}, \phi_E^h \boldsymbol{r}_E^h\right)-\left(\nabla \boldsymbol{u}^h, \nabla\left(\phi_E^h \boldsymbol{r}_E^h\right)\right)+\left(\nabla \cdot\left(\phi_E^h \boldsymbol{r}_E^h\right), p^h\right) \\
&=\sum_{K \in \omega_E}\left[\left(\boldsymbol{r}_K^h, \phi_E^h \boldsymbol{r}_E^h\right)_K+\left(\boldsymbol{f}-P^h \boldsymbol{f}, \phi_E^h \boldsymbol{r}_E^h\right)_K\right]+\left(\boldsymbol{r}_E^h, \phi_E^h \boldsymbol{r}_E^h\right)_E .
\end{aligned}$$

This identity is inserted into (5.72). With the Cauchy–Schwarz inequality (A.8), estimate (2.39), and (5.68), one obtains

$$\begin{aligned}
&\left\|\boldsymbol{r}_E^h\right\|_{L^2(E)}^2 \\
&\leq C\Big(\left\|\nabla(\boldsymbol{u}-\boldsymbol{u}^h)\right\|_{L^2(\omega_E)}\left\|\nabla(\phi_E^h \boldsymbol{r}_E^h)\right\|_{L^2(\omega_E)}+\left\|p-p^h\right\|_{L^2(\omega_E)}\left\|\nabla(\phi_E^h \boldsymbol{r}_E^h)\right\|_{L^2(\omega_E)} \\
&\quad+\left\|\boldsymbol{r}_K^h\right\|_{L^2(\omega_E)}\left\|\phi_E^h \boldsymbol{r}_E^h\right\|_{L^2(\omega_E)}+\left\|\boldsymbol{f}-P^h \boldsymbol{f}\right\|_{L^2(\omega_E)}\left\|\phi_E^h \boldsymbol{r}_E^h\right\|_{L^2(\omega_E)}\Big) \\
&\leq C\Big(h_E^{-1/2}\left\|\nabla(\boldsymbol{u}-\boldsymbol{u}^h)\right\|_{L^2(\omega_E)}+h_E^{-1/2}\left\|p-p^h\right\|_{L^2(\omega_E)}+h_E^{1/2}\left\|\boldsymbol{r}_K^h\right\|_{L^2(\omega_E)} \\
&\quad+h_E^{1/2}\left\|\boldsymbol{f}-P^h \boldsymbol{f}\right\|_{L^2(\omega_E)}\Big)\left\|\boldsymbol{r}_E^h\right\|_{L^2(\omega_E)} .
\end{aligned}$$

From the quasi-uniformity of the triangulation it follows that for each mesh cell K its diameter h_K can be estimated from below and above by a constant times h_E, where the constant is independent of the triangulation, the concrete mesh cells, and of the edges. Using this equivalence, dividing by $h_E^{-1/2}\left\|\boldsymbol{r}_E^h\right\|_{L^2(\omega_E)}$, and inserting estimate (5.71) gives the estimate for the face residual. ■

Theorem 5.52. Local lower, residual-based, a posteriori error estimate for conforming inf-sup stable finite element spaces. *Let the assumptions of Theorem 5.46 be satisfied and let*

$$\omega_K=\bigcup_{K' \in \mathcal{T}^h, \mathcal{E}(K') \subset \mathcal{E}(K) \neq \emptyset} K',$$

then there holds the estimate

$$\eta_K \leq C\left(\left\|\nabla(\boldsymbol{u}-\boldsymbol{u}^h)\right\|_{L^2(\omega_K)}+\left\|p-p^h\right\|_{L^2(\omega_K)}+h_K\left\|\boldsymbol{f}-P^h \boldsymbol{f}\right\|_{L^2(\omega_K)}\right), \tag{5.73}$$

where the constant is independent of the mesh width and of the solution of the Stokes equations.

Proof. It is with (2.38)

$$\left\|\nabla \cdot \boldsymbol{u}^h\right\|_{L^2(K)}=\left\|\nabla \cdot\left(\boldsymbol{u}-\boldsymbol{u}^h\right)\right\|_{L^2(K)} \leq \sqrt{d}\left\|\nabla\left(\boldsymbol{u}-\boldsymbol{u}^h\right)\right\|_{L^2(K)} .$$

With this estimate, estimate (5.69), and estimate (5.71) for all faces of K, (5.73) follows immediately. ■

Remark 5.53. On the local lower estimate (5.73).

- The use of residual-based error estimators for controlling an adaptive mesh refinement is popular in academia. Besides theoretical support, like estimates (5.73), the rather easy implementation of such estimators is a main reason.
- The constant in estimate (5.73) comes from the estimates for the local bubble functions given in Lemma 5.49.

□

5.4.2 The Dual Weighted Resdidual (DWR) Method

Remark 5.54. Motivation. A posteriori error estimates, like the residual-based estimate (5.63) for the Stokes equations, posses two drawbacks. First, such error estimates can be derived usually only for norms which are in some sense natural in the setup of the problem, like the norm in V or the $L^2(\Omega)$ norm of the velocity or pressure. The error in such norms is generally not of much interest in applications. There, rather errors of drag and lift coefficients or other quantities of interest are important. And second, a posteriori error estimates of type (5.63) have still an unknown factor on the right-hand side, like C in (5.63). This factor contains in particular contributions from the stability of the problem and of local interpolation error estimates, compare Remark 5.47. The interpolation error depends of course on the finite element space, which in turn depends on the underlying grid. The stability of the problem usually depends on coefficients of the problem and this dependency might be severe, like for convection-dominated convection-diffusion equations, see John (2000) for numerical studies which reveal such dependencies. All these unknown dependencies might lead to constants which differ much from 1. In such cases, the knowledge of only the computable factor on the right-hand side of residual-based a posteriori error estimates is solely of limited use, since it does not allow to draw reliable conclusions on the actual size of the error.

The dual weighted residual (DWR) method is an approach which deals with both drawbacks at the same time. It is a widely applicable approach which leads to error estimates for functionals of interest. In a first step, an abstract representation of the error for a functional of interest is derived, see Lemma 5.57. Then, the abstract framework is applied to variational problems, which leads to an error representation that involves a primal residual which involves the finite element solution, a dual residual which involves the solution of the discretized dual of a linearized problem, and a remainder, see Theorem 5.85. The remainder is considered to be of higher order and the arguments for the evaluation of the residuals will be approximated. Despite the approximations applied in this methodology, the experience in practice is that the obtained estimates are usually close to the errors.

The DWR method was proposed in Becker and Rannacher (1996, 1998), reviews as well as references to previous papers considering a posteriori estimates for functionals of interest can be found in Becker and Rannacher (2001); Bangerth and Rannacher (2003). □

Remark 5.55. Abstract setting. First, a general paradigm of a posteriori error analysis which is based on duality will be discussed. To this end consider a differentiable functional $L(\cdot)$ defined on some linear space X. A stationary point of this functional is a point $x \in X$ for which

$$L'(x)(y) = 0 \quad \forall\, y \in X, \tag{5.74}$$

where the prime refers to the first argument. The Galerkin approximation of this problem reads as follows: Find $x^h \in X^h \subset X$ such that

$$L'\left(x^h\right)\left(y^h\right) = 0 \quad \forall\, y^h \in X^h. \tag{5.75}$$

The second argument of the derivative and of all higher order derivatives is linear. □

Example 5.56. Energie functional. Let X be a Hilbert space whose inner product is given by the bilinear form $a(\cdot,\cdot)$. Then the energy functional is defined by

$$L(x) = \frac{1}{2}a(x,x) - f(x) \quad x \in X,$$

where $f(\cdot)$ is a continuous, linear functional on X. Consider the function

$$\Phi(\varepsilon) = L(x + \varepsilon y) \quad \forall\, y \in X.$$

Then, $x \in X$ is a stationary point if

$$\Phi'(0) = 0 \quad \Longleftrightarrow \quad L'(x)(y) = 0 \quad \forall\, y \in X.$$

A straightforward calculation, using the linearity of $f(\cdot)$, gives

$$\begin{aligned}
\Phi'(0) &= \lim_{\varepsilon\to 0} \frac{\Phi(\varepsilon) - \Phi(0)}{\varepsilon} \\
&= \lim_{\varepsilon\to 0} \frac{\frac{1}{2}a(x,x) + \varepsilon a(x,y) - \frac{\varepsilon^2}{2}a(y,y) - f(x) - \varepsilon f(y) - \frac{1}{2}a(x,x) + f(x)}{\varepsilon} \\
&= \lim_{\varepsilon\to 0} \left(a(x,y) - f(y) - \frac{\varepsilon}{2}a(y,y)\right) \\
&= a(x,y) - f(y) \quad \forall\, y \in X.
\end{aligned}$$

Thus, a stationary point has to satisfy the equation

$$a(x,y) = f(y) \quad \forall\, y \in X \quad \Longleftrightarrow \quad L'(x)(y) = 0 \quad \forall\, y \in X.$$

Computing the second variation, one finds that the solution of the equation for a stationary point x is a minimizer of the energy functional. □

Lemma 5.57. Abstract error representation. *Let the functional $L(\cdot)$ be three times differentiable. Then, for any solutions x of (5.74) and x^h of (5.75) there holds the error representation*

$$L(x) - L\left(x^h\right) = \frac{1}{2}L'\left(x^h\right)\left(x - I^h x\right) + \mathcal{R}, \tag{5.76}$$

where $I^h x \in X^h$ is arbitrary and with $e = x - x^h$ the remainder is given by

$$\mathcal{R} = \frac{1}{2}\int_0^1 \tau(\tau-1)L'''\left(x^h + \tau e\right)(e,e,e)\ d\tau. \tag{5.77}$$

Proof. With the fundamental theorem of calculus, one obtains

$$\begin{aligned}\int_0^1 L'\left(x^h + \tau e\right)(e)\ d\tau &= L\left(x^h + \tau e\right)\Big|_{\tau=0}^{\tau=1} = L\left(x^h + e\right) - L\left(x^h\right)\\ &= L(x) - L\left(x^h\right).\end{aligned} \tag{5.78}$$

In the next step, the integral will be represented with the help of the trapezoidal rule. To motivate this representation, let f be a sufficiently smooth function in $[t_0, t_1]$. Then, the trunctation error in the trapezoidal rule is given by

$$\int_{t_0}^{t_1} f(\tau)\ d\tau - \frac{f(t_1) + f(t_0)}{2}(t_1 - t_0).$$

Applying twice integration by parts and denoting $\bar{t} = (t_0 + t_1)/2$ yields

$$\begin{aligned}\frac{1}{2}\int_{t_0}^{t_1}\left(\left(\tau - \bar{t}\right)^2 - \left(\frac{t_1 - t_0}{2}\right)^2\right) f''(\tau)\ d\tau &= \int_{t_0}^{t_1}\left(\bar{t} - \tau\right) f'(\tau)\ d\tau\\ &= \int_{t_0}^{t_1} f(\tau)\ d\tau - \frac{f(t_1) + f(t_0)}{2}(t_1 - t_0).\end{aligned}$$

Applying the trapezoidal rule to the left-hand side of (5.78) and using the derived form of the truncation error, i.e., setting $t_1 = 1$, $t_0 = 0$, gives

$$\int_0^1 L'\left(x^h + \tau e\right)(e)\ d\tau = \frac{L'(x)(e) + L'\left(x^h\right)(e)}{2} + \frac{1}{2}\int_0^1 \tau(\tau-1)L'''\left(x^h + \tau e\right)(e,e,e)\ d\tau.$$

Since $e \in X$, one gets from (5.74) that $L'(x)(e) = 0$. Finally, the linearity of the second argument of the derivative and (5.75) leads for any $I^h x \in X^h$ to

$$L'\left(x^h\right)(e) = L'\left(x^h\right)\left(x - I^h x\right) + L'\left(x^h\right)\left(I^h x - x^h\right) = L'\left(x^h\right)\left(x - I^h x\right),$$

which completes the proof of the lemma. ■

Remark 5.58. Interpretation of the representation (5.76). The identity (5.76) shows that the error for the functional can be represented with the operator and the solution of the Galerkin problem (5.75), and a remainder which is cubic in the error. Note that the operator of the Galerkin problem (5.75) is linear. However, the test function of the Galerkin operator contains the solution of the continuous problem (5.74), which is also a linear problem. □

Remark 5.59. Application of the basic approach to a variational equation. Let V be a linear space and consider the equation: Find $u \in V$ such that

$$a(u;v) = a(u)(v) = f(v) \quad \forall\, v \in V, \tag{5.79}$$

where $a(\cdot)(\cdot) \;:\; V \times V \to \mathbb{R}$ is a differentiable form which is linear in the second argument and $f(\cdot) \;:\; V \to \mathbb{R}$ is a continuous linear functional. Let $V^h \subset V$ be a subspace, then the Galerkin approximation of (5.79) reads as follows: Find $u^h \in V^h$ such that

$$a\left(u^h\right)(v^h) = f\left(v^h\right) \quad \forall\, v^h \in V^h. \tag{5.80}$$

Denote by $J(\cdot) \;:\; V \to \mathbb{R}$ the functional of interest whose error $J(u) - J\left(u^h\right)$ should be minimized. To embed this problem into the general framework derived so far, one considers the optimization problem

$$J(u) \to \min, \quad a(u)(v) = f(v) \quad \forall\, v \in V.$$

If (5.79) has a unique solution, then the optimization problem is trivial since in this case there is just one argument for the functional. This methodology can be also applied for problems with non-unique solutions, like eigenvalue problems.

Next, the Euler–Lagrange approach for deriving conditions for the solution of the optimization problem is applied. To this end, one considers the Lagrangian functional

$$\mathcal{L}(u,z) = J(u) + f(z) - a(u)(z),$$

where $z \in V$ is called adjoint variable. A necessary condition for a minimizer is that it is a stationary point, i.e., there hold

$$\begin{aligned} 0 &= \partial_u \mathcal{L}(u,z) = J'(u)(w) - a'(u)(w,z) \quad \forall\, w \in V, \\ 0 &= \partial_z \mathcal{L}(u,z) = f(v) - a(u)(v) \quad \forall\, v \in V. \end{aligned} \tag{5.81}$$

In the second relation, the linearity of $f(\cdot)$ was used such that the second condition is just (5.79). Equation (5.81) is called the dual problem associated to the functional $J(\cdot)$. Again, one considers the Galerkin approximations of the two conditions: Find $\left(u^h, z^h\right) \in V^h \times V^h$ such that

$$\begin{aligned} a'\left(u^h\right)(w^h, z^h) &= J'\left(u^h\right)\left(w^h\right) \quad && \forall\, w^h \in V^h, \\ a\left(u^h\right)(v^h) &= f\left(v^h\right) && \forall\, v^h \in V^h. \end{aligned} \tag{5.82}$$

With the Galerkin solution, one defines the so-called primal residual

$$\rho \;:\; V \to V', \quad \rho(\cdot) = f(\cdot) - a\left(u^h\right)(\cdot)\,, \tag{5.83}$$

and the dual residual

$$\rho^* \ : \ V \to V', \quad J'\left(u^h\right)(\cdot) - a'\left(u^h\right)\left(\cdot, z^h\right). \tag{5.84}$$

□

Theorem 5.60. Error representation. *Let the form $a(\cdot)(\cdot)$ and the functional $J(\cdot)$ be three times differentiable, let (u, z) be any solution of (5.81) and let $\left(u^h, z^h\right)$ be any solution of (5.82). Then there is the error representation*

$$J(u) - J\left(u^h\right) = \frac{1}{2}\rho\left(z - I^h z\right) + \frac{1}{2}\rho^*\left(u - I^h u\right) + \mathcal{R}_a, \tag{5.85}$$

where $I^h z, I^h u \in V^h$ are arbitrary functions. Denoting $e = u - u^h$ and $e^* = z - z^h$, the remainder is given by

$$\begin{aligned}\mathcal{R}_a = \frac{1}{2}\int_0^1 \tau(\tau-1)\Big[&J'''\left(u^h + \tau e\right) - a'''\left(u^h + \tau e\right)\left(e, e, e, z^h + \tau e^*\right)\\ &-3a''\left(u^h + \tau e\right)\left(e, e, e^*\right)\Big]\, d\tau.\end{aligned}$$

Proof. This situation will be embedded into the general framework of Lemma 5.57. To this end, one sets $X = V \times V$, $X^h = V^h \times V^h$, $x = (u, z)$, $x^h = \left(u^h, z^h\right)$, and $L(x) = \mathcal{L}(u, z)$. Then, using the definition of the Lagrangian functional, the total derivative at x^h is given by

$$\begin{aligned}L'\left(x^h\right)(v_u, v_z) &= \partial_u \mathcal{L}\left(u^h, z^h\right)(v_u) + \partial_z \mathcal{L}\left(u^h, z^h\right)(v_z)\\ &= J'\left(u^h\right)(v_u) - a'\left(u^h\right)\left(v_u, z^h\right) + f(v_z) - a\left(u^h\right)(v_z)\end{aligned} \tag{5.86}$$

for arbitrary $(v_u, v_z) \in X$. Again, the linearity of $f(\cdot)$ and the second argument of $a(\cdot)(\cdot)$ was used. The application of (5.81), (5.82), (5.76), (5.86), (5.83), and (5.84) yields

$$\begin{aligned}&J(u) - J\left(u^h\right)\\ &= J(u) + f(z) - a(u)(z) - \left(J\left(u^h\right) + f\left(z^h\right) - a\left(u^h\right)\left(z^h\right)\right)\\ &= L(x) - L\left(x^h\right)\\ &= \frac{1}{2}L'\left(x^h\right)\left(x - I^h x\right) + \mathcal{R}\\ &= \frac{1}{2}\left(J'\left(u^h\right)\left(u - I^h u\right) - a'\left(u^h\right)\left(u - I^h u, z^h\right) + f\left(z - I^h z\right) - a\left(u^h\right)\left(z - I^h z\right)\right) + \mathcal{R}\\ &= \frac{1}{2}\left(\rho\left(z - I^h z\right) + \rho^*\left(u - I^h u\right)\right) + \mathcal{R},\end{aligned}$$

with arbitrary $\left(I^h u, I^h z\right) \in X^h$. The general form (5.77) of the remainder requires to compute the third derivative of $L(x) = \mathcal{L}(u, z)$. Formally, one has

$$L''' = \partial_{uuu}\mathcal{L} + 3\partial_{uuz}\mathcal{L} + \partial_{uzz}\mathcal{L} + \partial_{zzz}\mathcal{L}.$$

Since $\mathcal{L}$ depends only linearly on z, the last two terms vanish. The derivative of the other two terms at the required point $\left(x^h + \tau\left(e, e^*\right)\right)\left(\left(e, e^*\right), \left(e, e^*\right), \left(e, e^*\right)\right)$ gives just the terms in $\mathcal{R}_a$

$$
\begin{aligned}
& L''' \left(x^h + \tau \left(e, e^*\right)\right) \left(\left(e, e^*\right), \left(e, e^*\right), \left(e, e^*\right)\right) \\
& \quad = J''' \left(u^h + \tau e\right) - a''' \left(u^h + \tau e\right) \left(e, e, e, z^h + \tau e^*\right) - 3a'' \left(u^h + \tau e\right) \left(e, e, e^*\right). \quad (5.87)
\end{aligned}
$$

For the second term on the left-hand side, the three derivatives with respect to u lead to three times the argument e and for the last term, the two derivatives with respect to u and the last derivative with respect to z lead to twice the argument e and once the argument e^*. Note that the prime refers only to the first argument of $a(\cdot)(\cdot)$, which gives the last term in (5.87). ■

Remark 5.61. Linear variational problem and linear functional. Consider a linear variational problem

$$a(u)(v) = a(u, v) = f(v) \quad \forall\, v \in V$$

and the corresponding Galerkin approximation

$$a\left(u^h, v^h\right) = f\left(v^h\right) \quad \forall\, v^h \in V^h.$$

Subtracting both equations, one gets the Galerkin orthogonality

$$a\left(u - u^h, v^h\right) = 0 \quad \forall\, v^h \in V^h.$$

Since $a'\left(u^h\right)\left(\cdot, z^h\right) = a\left(\cdot, z^h\right)$ and $J'\left(u^h\right)(\cdot) = J(\cdot)$, one obtains with the Galerkin orthogonality, since $z^h \in V^h$, the linearity of the functional, and $I^h u = u^h$

$$\rho^*\left(u - I^h u\right) = J\left(u\right) - a\left(u, z^h\right) - J\left(u^h\right) + a\left(u^h, z^h\right) = J\left(u - u^h\right) = J(e).$$

Inserting this expression into the error representation (5.85) and observing that the remainder vanishes, since all higher order derivatives of the variational form and the functional vanish, yields

$$J(u) - J\left(u^h\right) = J(e) = \rho\left(z - I^h z\right).$$

Thus, for this special case one needs to compute only the primal residual, evaluated for the difference of the solution of the dual problem and an arbitrary interpolation. □

Remark 5.62. Dual linearized problem for the steady-state Navier–Stokes equations. Consider the Navier–Stokes equations in V_{div} and let $\boldsymbol{w} \in V_{\text{div}}$ be an arbitrary element. Then, the linearization of the Navier–Stokes equations at $\boldsymbol{w}$ is given by the left-hand side of (5.55), which reads as follows

$$a(\boldsymbol{u}, \boldsymbol{v}) + n\left(\boldsymbol{w}, \boldsymbol{u}, \boldsymbol{v}\right) + n\left(\boldsymbol{u}, \boldsymbol{w}, \boldsymbol{v}\right) \quad \forall\, \boldsymbol{v} \in V_{\text{div}}. \tag{5.88}$$

Here, $\boldsymbol{u} \in V_{\text{div}}$ is the solution of the Navier–Stokes equations. For the dual problem, ansatz and test functions change their role, such that the left-hand side of the dual problem is

$$a(\boldsymbol{v},\boldsymbol{\phi}) + n(\boldsymbol{w},\boldsymbol{v},\boldsymbol{\phi}) + n(\boldsymbol{v},\boldsymbol{w},\boldsymbol{\phi}) \quad \forall\, \boldsymbol{v} \in V_{\mathrm{div}}, \tag{5.89}$$

where $\boldsymbol{\phi} \in V_{\mathrm{div}}$ is the solution of the dual problem for a given right-hand side. For illustrating which type of problem one has to solve as dual problem, it is of advantage to write the forms in the usual form, i.e., such that the test function is in the last argument. Since the viscous term is symmetric, it is

$$a(\boldsymbol{v},\boldsymbol{\phi}) = a(\boldsymbol{\phi},\boldsymbol{v}).$$

For the second term of (5.89), one obtains with the product rule, see (1.24), $\boldsymbol{w} \in V_{\mathrm{div}}$, integration by parts, and a direct calculation

$$\begin{aligned} n(\boldsymbol{w},\boldsymbol{v},\boldsymbol{\phi}) &= ((\boldsymbol{w}\cdot\nabla)\boldsymbol{v},\boldsymbol{\phi}) = \left(\nabla\cdot\left(\boldsymbol{v}\boldsymbol{w}^T\right),\boldsymbol{\phi}\right) - ((\nabla\cdot\boldsymbol{w})\boldsymbol{v},\boldsymbol{\phi}) \\ &= -\left(\boldsymbol{v}\boldsymbol{w}^T,\nabla\boldsymbol{\phi}\right) = -((\boldsymbol{w}\cdot\nabla)\boldsymbol{\phi},\boldsymbol{v}) = -n(\boldsymbol{w},\boldsymbol{\phi},\boldsymbol{v}), \end{aligned}$$

where the direct calculation gives

$$\begin{aligned} &\left(\boldsymbol{v}\boldsymbol{w}^T,\nabla\boldsymbol{\phi}\right) \\ &= \left(\begin{pmatrix} v_1w_1 & v_1w_2 & v_1w_3 \\ v_2w_1 & v_2w_2 & v_2w_3 \\ v_3w_1 & v_3w_2 & v_3w_3 \end{pmatrix}, \begin{pmatrix} \partial_x\phi_1 & \partial_y\phi_1 & \partial_z\phi_1 \\ \partial_x\phi_2 & \partial_y\phi_2 & \partial_z\phi_2 \\ \partial_x\phi_3 & \partial_y\phi_3 & \partial_z\phi_3 \end{pmatrix} \right) \\ &= \int_\Omega \Big[v_1w_1\partial_x\phi_1 + v_1w_2\partial_y\phi_1 + v_1w_3\partial_z\phi_1 + v_2w_1\partial_x\phi_2 + v_2w_2\partial_y\phi_2 \\ &\quad + v_2w_3\partial_z\phi_2 + v_3w_1\partial_x\phi_3 + v_3w_2\partial_y\phi_3 + v_3w_3\partial_z\phi_3 \Big]\, d\boldsymbol{x}, \end{aligned}$$

and

$$\begin{aligned} &((\boldsymbol{w}\cdot\nabla)\boldsymbol{\phi},\boldsymbol{v}) \\ &= \left(\begin{pmatrix} w_1\partial_x\phi_1 + w_2\partial_y\phi_1 + w_3\partial_z\phi_1 \\ w_1\partial_x\phi_2 + w_2\partial_y\phi_2 + w_3\partial_z\phi_2 \\ w_1\partial_x\phi_3 + w_2\partial_y\phi_3 + w_3\partial_z\phi_3 \end{pmatrix}, \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} \right) \\ &= \int_\Omega \Big[v_1w_1\partial_x\phi_1 + v_1w_2\partial_y\phi_1 + v_1w_3\partial_z\phi_1 + v_2w_1\partial_x\phi_2 + v_2w_2\partial_y\phi_2 \\ &\quad + v_2w_3\partial_z\phi_2 + v_3w_1\partial_x\phi_3 + v_3w_2\partial_y\phi_3 + v_3w_3\partial_z\phi_3 \Big]\, d\boldsymbol{x}. \end{aligned}$$

For the last term of (5.89), a direct calculation yields

$$n(\boldsymbol{v},\boldsymbol{w},\boldsymbol{\phi}) = \left((\nabla\boldsymbol{w})^T\boldsymbol{\phi},\boldsymbol{v}\right),$$

where the details of this calculation are as follows

$$
\begin{aligned}
& n\left(\boldsymbol{v}, \boldsymbol{w}, \boldsymbol{\phi}\right) \\
& = \left(\left(\boldsymbol{v} \cdot \nabla\right) \boldsymbol{w}, \boldsymbol{\phi}\right) \\
& \quad \int_\Omega \Big[v_1 \partial_x w_1 \phi_1 + v_2 \partial_y w_1 \phi_1 + v_3 \partial_z w_1 \phi_1 + v_1 \partial_x w_2 \phi_2 + v_2 \partial_y w_2 \phi_2 \\
& \quad + v_3 \partial_z w_2 \phi_2 + v_1 \partial_x w_3 \phi_3 + v_2 \partial_y w_3 \phi_3 + v_3 \partial_z w_3 \phi_3\Big] \, d\boldsymbol{x}
\end{aligned}
$$

and

$$
\begin{aligned}
& \left(\left(\nabla \boldsymbol{w}\right)^T \boldsymbol{\phi}, \boldsymbol{v}\right) \\
& = \left(\begin{pmatrix} \partial_x w_1 \phi_1 + \partial_x w_2 \phi_2 + \partial_x w_3 \phi_3 \\ \partial_y w_1 \phi_1 + \partial_y w_2 \phi_2 + \partial_y w_3 \phi_3 \\ \partial_z w_1 \phi_1 + \partial_z w_2 \phi_2 + \partial_z w_3 \phi_3 \end{pmatrix}, \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix}\right) \\
& = \int_\Omega \Big[v_1 \partial_x w_1 \phi_1 + v_1 \partial_x w_2 \phi_2 + v_1 \partial_x w_3 \phi_3 + v_2 \partial_y w_1 \phi_1 + v_2 \partial_y w_2 \phi_2 \\
& \quad + v_2 \partial_y w_3 \phi_3 + v_3 \partial_z w_1 \phi_1 + v_3 \partial_z w_2 \phi_2 + v_3 \partial_z w_3 \phi_3\Big] \, d\boldsymbol{x}.
\end{aligned}
$$

Summerizing the terms, one finds that the left-hand side of the dual linearized problem (5.89) can be written in the form

$$a(\boldsymbol{\phi}, \boldsymbol{v}) - n\left(\boldsymbol{w}, \boldsymbol{\phi}, \boldsymbol{v}\right) + \left(\left(\nabla \boldsymbol{w}\right)^T \boldsymbol{\phi}, \boldsymbol{v}\right), \tag{5.90}$$

which corresponds to the strong formulation

$$-\nu \Delta \boldsymbol{\phi} - (\boldsymbol{w} \cdot \nabla)\boldsymbol{\phi} + \left(\nabla \boldsymbol{w}\right)^T \boldsymbol{\phi}.$$

Analogously to the linearized Navier–Stokes problem (5.88) there is a viscous term, a convective term, and a reactive term. However the convection field of the dual problem (5.90) is the negative of the convection field of the linearized Navier–Stokes problem (5.88). □

Remark 5.63. Practical aspects.

- The computation of the residuals on the right-hand side of (5.85) requires the knowledge of u, z, u^h, and z^h. The solutions of the discrete problems appear in the definition of the residuals (5.83) and (5.84). In practice, it will not be possible to solve the continuous problems for obtaining u und z. However, having computed u^h, the discretized dual problem (5.82)

  $$a'\left(u^h\right)(w^h, z^h) = J'\left(u^h\right)\left(w^h\right) \quad \forall\, w^h \in V^h,$$

  can be solved such that u^h and z^h are available.
  If a coarse grid is available, then one can apply a local post-processing, i.e., a patchwise interpolation into a higher order finite element space on the coarser grid to compute functions $I_{\text{ho}}^{2h} u^h$ and $I_{\text{ho}}^{2h} z^h$. With these functions, one approximates in (5.85)

$$\left(z - I^h z\right)|_K \approx \left(\tilde{z} - I^h \tilde{z}\right)|_K := \left(I_{\mathrm{ho}}^{2h} z^h - z^h\right)|_K, \quad \forall\, K \in \mathcal{T}^h,$$
$$\left(u - I^h u\right)|_K \approx \left(\tilde{u} - I^h \tilde{u}\right)|_K := \left(I_{\mathrm{ho}}^{2h} u^h - u^h\right)|_K, \quad \forall\, K \in \mathcal{T}^h.$$

references for NSE and alternative ways

- The remainder $\mathcal{R}_a$ in (5.85) is usually neglected.
- The dual problem is linear, i.e., for nonlinear problems like the stationary Navier–Stokes equations, the costs of solving the dual problem are of the order of performing one step in their iterative solution.

some citations with applications □