

# Finite Element Methods for the Incompressible Navier-Stokes Equations

Rolf Rannacher \*

Institute of Applied Mathematics  
University of Heidelberg  
INF 293/294, D-69120 Heidelberg, Germany  
E-Mail: [rannacher@iwr.uni-heidelberg.de](mailto:rannacher@iwr.uni-heidelberg.de)  
URL: <http://gaia.iwr.uni-heidelberg.de>

Version of August 11, 1999

## Abstract

These notes are based on lectures given in a Short Course on Theoretical and Numerical Fluid Mechanics in Vancouver, British Columbia, Canada, July 27-28, 1996, and at several other places since then. They provide an introduction to recent developments in the numerical solution of the Navier-Stokes equations by the finite element method. The material is presented in eight sections:

1. Introduction: Computational aspects of laminar flows
2. Models of viscous flow
3. Spatial discretization by finite elements
4. Time discretization and linearization
5. Solution of algebraic systems
6. A review of theoretical analysis
7. Error control and mesh adaptation
8. Extension to weakly compressible flows

Theoretical analysis is offered to support the construction of numerical methods, and often computational examples are used to illustrate theoretical results. The variational setting of the finite element Galerkin method provides the theoretical framework. The goal is to guide the development of more efficient and accurate numerical tools for computing viscous flows. A number of open theoretical problems will be formulated, and many references are made to the relevant literature.

---

\*The author acknowledges the support by the German Research Association (DFG) through the SFB 359 "Reactive Flow, Diffusion and Transport" at the University of Heidelberg, Im Neuenheimer Feld 294, D-69120 Heidelberg, Germany.

# 1 Introduction

In the following sections, we will discuss a computational methodology for simulating viscous incompressible laminar flows. The description of the numerical algorithms will be accompanied by a theoretical analysis so far as it is relevant to understanding the performance of the method. In this sense, these notes are meant as a contribution of Mathematics to “CFD” (Computational Fluid Dynamics).

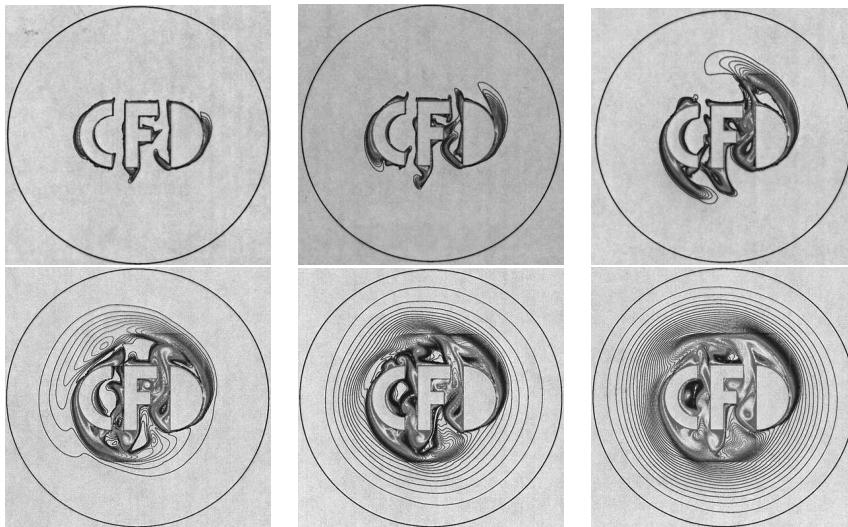


Figure 1: Nonstationary flow around “CFD” for  $Re = 500$ , driven by rotation of the outer circle and visualized by temperature isolines; from Turek [98].

The established model for viscous Newtonian incompressible flow is the the system of Navier-Stokes equations,

$$\partial_t v - \nu \Delta v + v \cdot \nabla v + \nabla p = f, \quad \nabla \cdot v = 0, \quad (1)$$

in some region  $\Omega \times (0, T)$  with appropriate initial and boundary conditions. We concentrate on “laminar” flows, i.e., on flows with Reynolds number in the range  $1 \leq Re \leq 10^5$ , where  $Re \sim \bar{v}l/\nu$ . The numerical solution of this system involves several typical difficulties:

- Complicated flow structure  $\Rightarrow$  fine meshes!
- $Re \gg 1 \Rightarrow$  locally refined and anisotropic meshes in boundary layers!
- Dominant nonlinear effects  $\Rightarrow$  stability problems!
- Constraint  $\nabla \cdot v = 0 \Rightarrow$  implicit solution!
- Sensitive quantities  $\Rightarrow$  solution-adapted meshes!

Accurate flow prediction requires the use of large computer power, particularly for the extension from 2D to 3D, from stationary to nonstationary flows, and from qualitative results to quantitatively accurate results. The key goals in the developing tools for computing laminar flows are:

- *fast* (nonstationary calculations in minutes or hours),
- *cheap* (simulations on workstations),
- *flexible* (general purpose solver),
- *accurate* (adaptive error control).

### 1.1 Solution method

The method of choice in these notes is the “Finite Element Method” (FEM) for computing the primitive variables  $v$  (velocity) and  $p$  (pressure). This special Galerkin method is based on a variational formulation of the Navier-Stokes problem in appropriate function spaces, and determines “discrete” approximations in certain finite dimensional subspaces (“trial spaces”) consisting of piecewise polynomial functions. By this approach the discretization inherits most of the rich structure of the continuous problem, which, on the one hand provides a *high computational flexibility* and on the other hand facilitates a *rigorous mathematical error analysis*. These are the main aspects which make the FEM increasingly attractive in CFD. For completeness, we briefly comment on the essential features of the main competitors of the FEM:

- **Finite difference methods (FDM):** Approximation of the Navier-Stokes equations in their “strong” form by finite differences:
  - + easy implementation,
  - problems along curved boundaries,
  - difficult stability and convergence analysis,
  - mesh adaptation difficult.
- **Finite volume methods (FVM):** Approximation of the Navier-Stokes equations as a system of (cell-wise) conservation equations:
  - + based on “physical” conservation properties,
  - problems on unstructured meshes,
  - difficult stability and convergence analysis,
  - only heuristic mesh adaptation.
- **Spectral Galerkin methods:** Approximation of the Navier-Stokes equations in their variational form by a Galerkin method with “high-order” polynomial trial functions:
  - + high accuracy,
  - treatment of complex domains difficult,
  - mesh and order (hp)-adaptation difficult.

This brief classification must be superficial and is based on personal taste. The details are the subject of much controversial discussion concerning the pros and cons of the various methods and their variants. However, this conflict is partially resolved in many cases, as the differences between the methods, particularly between FEM and FVM, often disappear on general meshes. In fact, some of the FVMs can be interpreted as variants of certain “mixed” FEMs.

## 1.2 Examples of computable viscous flows

Below, we give some examples of flows which can be computed by the methods described in these notes. More examples including movies of nonstationary flows can be seen on our homepage: <http://gaia.iwr.uni-heidelberg.de/>. Some of the computer codes are available for research purposes:

- FEATFLOW Code (FORTRAN 77) by S. Turek [96], [97]:  
<http://gaia.iwr.uni-heidelberg.de/~featflow/>.
- deal.II Code (C++) by W. Bangerth and G. Kanschat [5]:  
<http://gaia.iwr.uni-heidelberg.de/~deal/>.

A collection of experimental photographs of such “computable” flows can be found in Van Dyke’s book *“An Album of Fluid Motion”* [99]. In the following, we present some examples of viscous flows which have been computed by the methodology described in these notes. Most of these results emerged as side products in the course of developing the numerical solvers and testing them for standard benchmark problems. All computations were done on normal work stations.

**Example 1. Cavity flow:** The first example is stationary and nonstationary flow in a cavity driven by flow along the upper boundary (“driven cavity”).

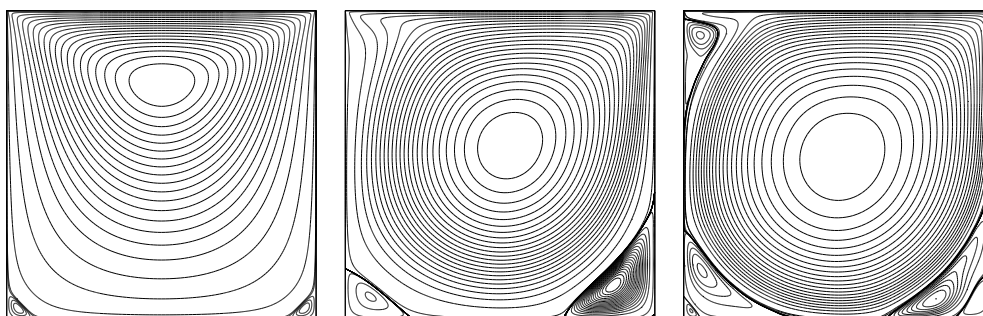


Figure 2: Stationary driven cavity flow in 2D for  $Re = 1, 1000, 9000$  (from left to right); for  $Re > 10000$  the flow becomes nonstationary; from Turek [98].

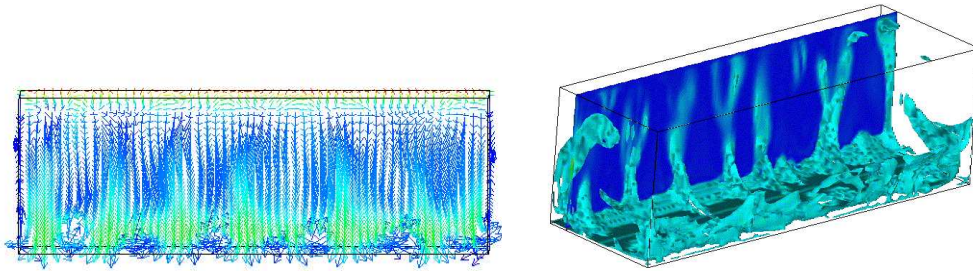


Figure 3: Simulation of nonstationary 3D driven cavity flow for  $Re = 100$ ; from Oswald [66].

**Example 2. Vortex shedding:** The second example is nonstationary flow around a circular cylinder (“von Kármán vortex street”)

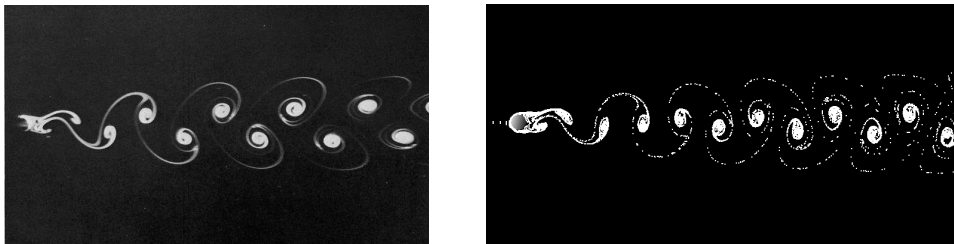


Figure 4: Von Kármán vortex street; experiment with  $Re = 105$  (left; from Van Dyke [99]), and 2D computation with  $Re = 100$  (right; from Turek [98]).

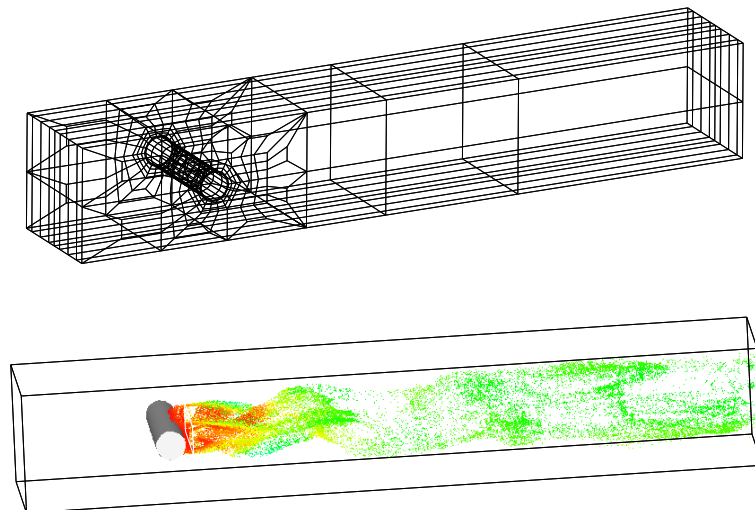


Figure 5: 3D simulation of vortex shedding behind a cylinder for  $Re = 100$ , coarse grid and flow visualized by particle tracing; from Oswald [66].

**Example 3. Leapfrogging of vortex rings:** Two successive puffs of fluid are injected through a hole and develop into vortex rings dancing around each other. The second ring travels faster in the induced wake of the first and slips through it. Then the first ring slips through the second, and so on.

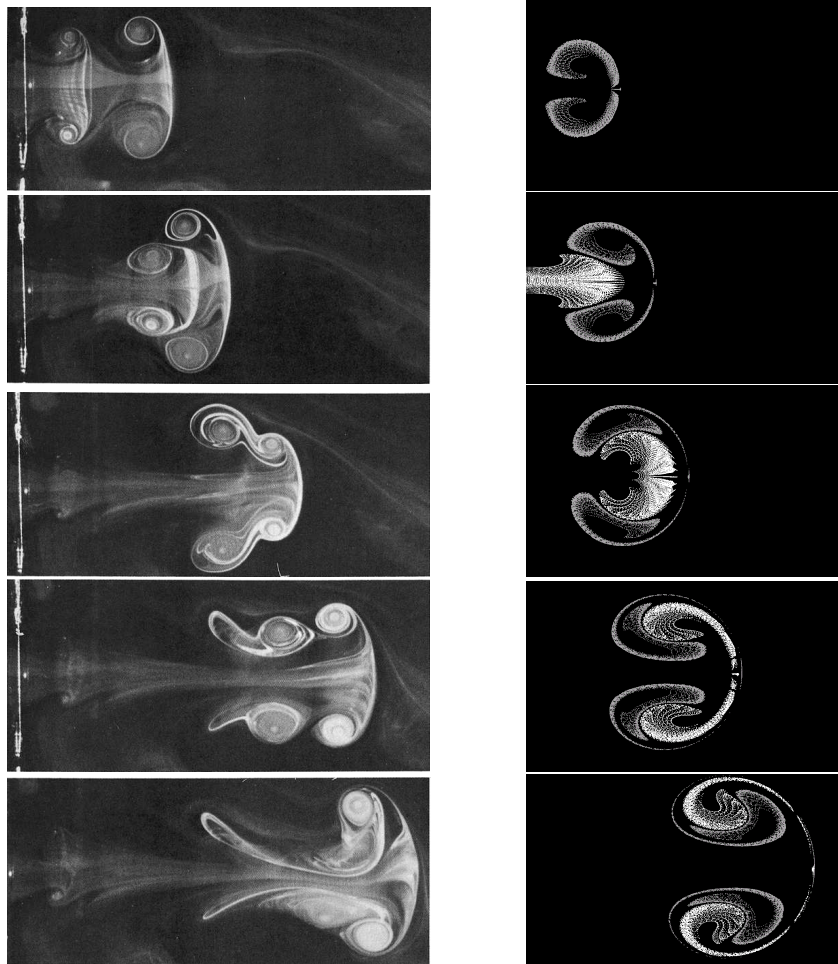


Figure 6: Leapfrogging of two vortex rings; experiment for  $Re \approx 1600$  (left; from Van Dyke [99]) and 2D computation for  $Re \approx 500$  (right; from [46]).

### 1.2.1 EXTENSIONS BEYOND STANDARD NAVIER-STOKES FLOW

The numerical methodology described in these notes has primarily been developed for computing viscous incompressible Newtonian flows. However, extensions are possible in several directions. These include flows in regions with moving boundaries, as for example pipe flow driven by rotating propellers, and flow of non-Newtonian fluids modelled by a simple power-law rule. The extension to certain low-speed compressible flows will be discussed in more detail below in Section 8.

### 1) Flow regions with moving boundaries

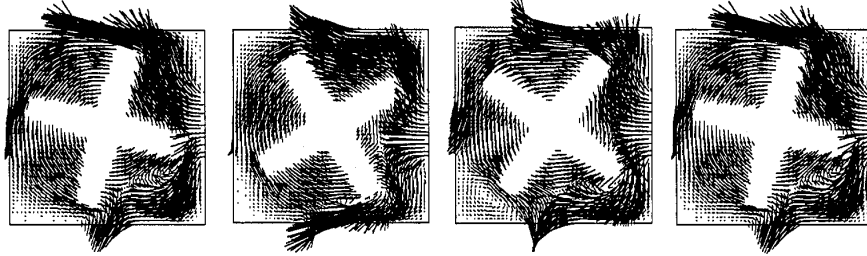


Figure 7: Velocity plot of 2D flow in a box driven by a rotating cross, computed by a “virtual boundary” technique; from Turek [97].

### 2) Flow of a non-Newtonian fluid

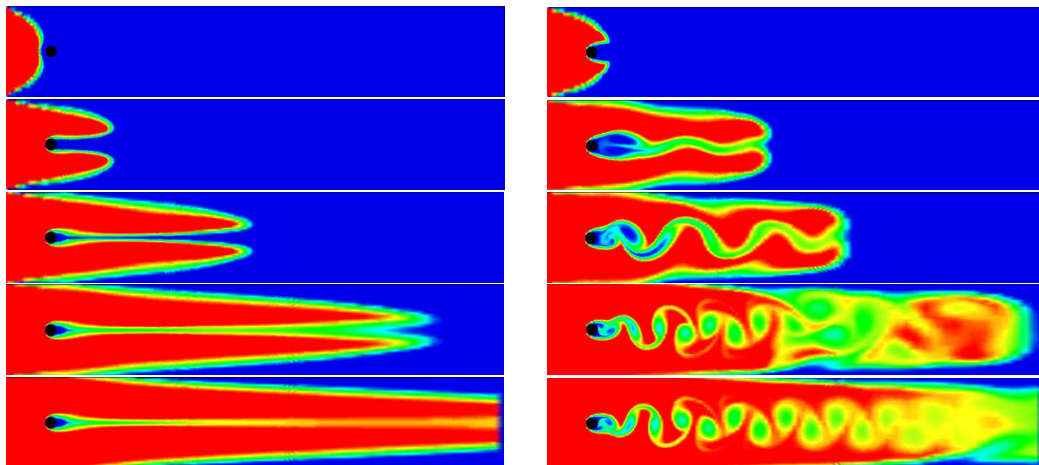


Figure 8: Computation of the flow of a non-Newtonian fluid around a circular obstacle in a 2D channel (“power-law”  $\nu = \nu(1 + |D(v)|)^{-1}$ ): stationary flow in the Newtonian case (left) and nonstationary flow in the non-Newtonian case (right), both for the same Reynolds number  $Re = 20$ ; from Turek [98].

### 3) Low-Mach-number compressible flow

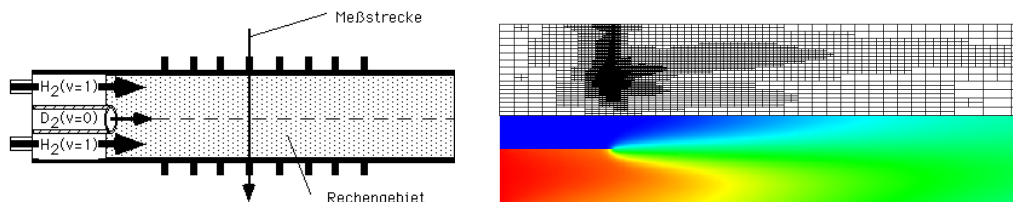


Figure 9: Computation of compressible flow in a chemical flow reactor: flow configuration and mass fraction of excited  $H_2^{(\nu=1)}$  computed on a locally refined mesh; from Waguët [103] and [10].

## 2 Models of viscous flow

The mathematical model for describing viscous (Newtonian) flows is the system of Navier–Stokes equations, which are the equations of conservation of mass, momentum and energy:

$$\partial_t \rho + \nabla \cdot [\rho v] = 0, \quad (1)$$

$$\partial_t(\rho v) + \rho v \cdot \nabla v - \nabla \cdot [\mu \nabla v + \frac{1}{3} \mu \nabla \cdot v I] + \nabla p_{tot} = \rho f, \quad (2)$$

$$\partial_t(c_p \rho T) + c_p \rho v \cdot \nabla T - \nabla \cdot [\lambda \nabla T] = h. \quad (3)$$

Here,  $v$  is the velocity,  $\rho$  the density,  $p_{tot}$  the (total) pressure, and  $T$  the temperature of the fluid occupying a two- or three-dimensional region  $\Omega$ . The parameters  $\mu > 0$  (dynamic viscosity),  $c_p > 0$  (heat capacity) and  $\lambda > 0$  (heat conductivity) characterize the properties of the fluid. The volume force  $f$  and the heat source  $h$  are explicitly given. Since we only consider low-speed flows, the influence of stress and hydrodynamic pressure in the energy equation can be neglected. In temperature-driven flows,  $h$  may implicitly depend on the temperature and further quantities describing heat release, as for example by chemical reactions. Such “weakly compressible” flows will be briefly considered at the end of these notes in Section 8. Here, the fluid is assumed to be *incompressible* and the density to be homogeneous,  $\rho \equiv \rho_0 = \text{const.}$ , so that (1) reduces to the incompressibility constraint

$$\nabla \cdot v = 0. \quad (4)$$

In this model, we consider as the primal unknowns the velocity  $v$ , the pressure  $p = p_{tot}$ , and the temperature  $T$ . For most parts of the discussion, the flow is assumed to be isothermal, so that the energy equation decouples from the momentum and continuity equations, and the temperature only enters through the viscosity parameter. The system is closed by imposing appropriate initial and boundary conditions for the flow variables

$$v|_{t=0} = v^0, \quad v|_{\Gamma_{rigid}} = 0, \quad v|_{\Gamma_{in}} = v^{in}, \quad (\mu \partial_n v + pn)|_{\Gamma_{out}} = 0, \quad (5)$$

and corresponding ones for the temperature, where  $\Gamma_{rigid}$ ,  $\Gamma_{in}$ ,  $\Gamma_{out}$  are the rigid part, the inflow part and the outflow part of the boundary  $\partial\Omega$ , respectively. The role of the natural outflow boundary condition on  $\Gamma_{out}$  will be explained in more detail below.

In the isothermal case, assuming for simplicity that  $\rho_0 = 1$ , the Navier–Stokes system can be written in short as

$$\partial_t v + v \cdot \nabla v - \nu \Delta v + \nabla p = f, \quad \nabla \cdot v = 0, \quad (6)$$

with the kinematic viscosity parameter  $\nu = \mu/\rho_0$ . In this formulation the domain  $\Omega$  may be taken two- or three-dimensional according to the particular requirements of the simulation. In our examples, we shall mostly refer to the



two-dimensional case. Through a scaling transformation this problem is made dimensionless, with the Reynolds Number  $Re = UL/\nu$  as the characteristic parameter, where  $U$  is the reference velocity (e.g.,  $U \approx \max |v^{\text{in}}|$ ) and  $L$  the characteristic length (e.g.,  $L \approx \text{diam}(\Omega)$ ), of the problem.

As common in the mathematical literature, we assume that  $U = 1$  and  $L = 1$  and consider  $\nu := 1/Re$  as a *dimensionless* parameter characterizing in some sense the “complexity” of the flow problem. Then, the length of the time interval over which the solution develops its characteristic features is  $T \approx 1/\nu$ , and the relevant scale of its spatial structures is  $\delta x \approx \sqrt{\nu}$  (width of boundary layers). This has to be kept in mind when the right spatial mesh size  $h$  and the time step  $k$  is chosen for a numerical simulation which is supposed to resolve all structures of the flow; for a more detailed discussion of the issue of reliable transient flow calculation see [54].

## 2.1 Variational formulation

The finite element discretization of the Navier-Stokes problem (6) is based on its variational formulation. To this end, we use the following sub-spaces of the usual Lebesgue function space  $L^2(\Omega)$  of square-integrable functions on  $\Omega$ :

$$\begin{aligned} L_0^2(\Omega) &= \{ \varphi \in L^2(\Omega) : (\varphi, 1) = 0 \}, \\ H^1(\Omega) &= \{ v \in L^2(\Omega), \partial_i v \in L^2(\Omega), 1 \leq i \leq d \}, \\ H_0^1(\Gamma; \Omega) &= \{ v \in H^1(\Omega), v|_\Gamma = 0 \}, \quad \Gamma \subset \partial\Omega, \end{aligned}$$

and the corresponding inner products and norms

$$\begin{aligned} (u, v) &= \int_\Omega uv \, dx, \quad \|v\| = (v, v)^{1/2}, \\ \|\nabla v\| &= (\nabla v, \nabla v)^{1/2}, \quad \|v\|_1 = (\|v\|^2 + \|\nabla v\|^2)^{1/2}. \end{aligned}$$

These are all spaces of  $\mathbb{R}$ -valued functions. Spaces of  $\mathbb{R}^d$ -valued functions  $v = (v_1, \dots, v_n)$  are denoted by boldface-type, but no distinction is made in the notation of norms and inner products; thus  $\mathbf{H}_0^1(\Gamma; \Omega) = H_0^1(\Gamma; \Omega)^d$  has norm  $\|v\|_1 = (\sum_{i=1}^d \|v_i\|_1^2)^{1/2}$ , etc. All the other notation is self-evident:  $\partial_t u = \partial u / \partial t$ ,  $\partial_i u = \partial u / \partial x_i$ ,  $\partial_n v = n \cdot \nabla v$ ,  $\partial_\tau = \tau \cdot \nabla v$  etc., where  $n$  and  $\tau$  are the normal and tangential unit vectors along the boundary  $\partial\Omega$ .

The pressure  $p$  in the Navier-Stokes equations is uniquely (possibly up to a constant) determined by the velocity field  $v$ . This follows from the fact that every bounded functional  $F(\cdot)$  on  $\mathbf{H}_0^1(\Gamma; \Omega)$  which vanishes on the subspace

$$\mathbf{J}_1(\Gamma; \Omega) = \{ v \in \mathbf{H}_0^1(\Gamma; \Omega), \nabla \cdot v = 0 \}$$

can be expressed in the form  $F(\varphi) = (p, \nabla \cdot \varphi)$  for some  $p \in L^2(\Omega)$ . Further, there holds the stability estimate (“inf-sup” stability)

$$\inf_{q \in L^2(\Omega)} \left\{ \sup_{\phi \in \mathbf{H}_0^1(\Gamma; \Omega)} \frac{(q, \nabla \cdot \phi)}{\|\nabla \phi\|} \right\} \geq \gamma_0 > 0, \quad (7)$$

where  $L^2(\Omega)$  has to be replaced by  $L_0^2(\Omega)$  in the case  $\Gamma = \partial\Omega$ . For proofs of these facts, one may consult the first parts of the books of Ladyshenskaya [58], Temam [88] and Girault/Raviart [29]. Finally, we introduce the notation

$$a(u, v) := \nu(\nabla u, \nabla v), \quad n(u, v, w) := (u \cdot \nabla v, w), \quad b(p, v) := -(p, \nabla \cdot v),$$

and the abbreviations

$$\mathbf{H} := \mathbf{H}_0^1(\Gamma; \Omega), \quad L := L^2(\Omega) \quad (L := L_0^2(\Omega) \text{ in the case } \Gamma = \partial\Omega),$$

where  $\Gamma = \Gamma_{in} \cup \Gamma_{rigid}$ . Then, the variational formulation of the Navier-Stokes problem (6), reads as follows: Find functions  $v(\cdot, t) \in v^{in} + \mathbf{H}$  and  $p(\cdot, t) \in L$ , such that  $v|_{t=0} = v^0$ , and setting  $\Gamma := \Gamma_{in} \cup \Gamma_{rigid}$ ,

$$(\partial_t v, \varphi) + a(v, \varphi) + n(v, v, \varphi) + b(p, v) = (f, \varphi) \quad \forall \varphi \in \mathbf{H}, \quad (8)$$

$$(\nabla \cdot v, \chi) = 0 \quad \forall \chi \in L. \quad (9)$$

It is well known that in two space dimensions the pure Dirichlet problem (8), (9), with  $\Gamma_{out} = \emptyset$ , possesses a unique solution on any time interval  $[0, T]$ , which is also a classical solution if the data of the problem are smooth enough. For small viscosity, i.e., large Reynolds numbers, this solution may be unstable. In three dimensions, the existence of a unique solution is known only for sufficiently small data, e.g.,  $\|v^0\|_1 \approx \nu$ , or on sufficiently short intervals of time,  $0 \leq t \leq T$ , with  $T \approx \nu$ .

## 2.2 Regularity of solution

We collect some results concerning the regularity of the variational solution of the Navier-Stokes problem which are relevant for the understanding of its numerical approximation. One obtains quantitative regularity bounds from the following sequence of differential identities

$$\begin{aligned} \frac{1}{2} d_t \|v\|^2 + \nu \|\nabla v\|^2 &= (f, v), \\ \frac{1}{2} d_t \|\nabla v\|^2 + \nu \|\Delta v\|^2 &= -(f, \Delta v) + (v \cdot \nabla v, \Delta v), \\ \frac{1}{2} d_t \|\partial_t v\|^2 + \nu \|\nabla \partial_t v\|^2 &= (\partial_t f, \partial_t v) - (\partial_t v \cdot \nabla v, \partial_t v), \\ \frac{1}{2} d_t \|\nabla \partial_t v\|^2 + \nu \|\Delta \partial_t v\|^2 &= -(\partial_t f, \Delta \partial_t v) + (\partial_t v \cdot \nabla v, \Delta \partial_t v) + \dots, \\ \frac{1}{2} d_t \|\partial_t^2 v\|^2 + \nu \|\nabla \partial_t^2 v\|^2 &= (\partial_t^2 f, \partial_t^2 v) - (\partial_t^2 v \cdot \nabla v, \partial_t^2 v) - \dots, \\ &\dots \end{aligned}$$

which are easily derived by standard energy arguments; see [40] and [41]. Assuming a bound on the Dirichlet norm of  $v$ ,

$$\sup_{t \in (0, T]} \|\nabla v(t)\| \leq M, \quad (10)$$

the above estimates together with the usual elliptic regularity results imply that  $v$  is smooth on  $\bar{\Omega}$  for  $0 < t \leq T$ , if all the data and  $\partial\Omega$  are

smooth. However, for the purposes of numerical analysis one needs regularity estimates which hold uniformly for  $t \rightarrow 0$ . To get such information from the above equations requires starting values for all the quantities  $\|v\|, \|\nabla v\|, \|\partial_t v\|, \|\nabla \partial_t v\|, \|\partial_t^2 v\|$ , etc., at  $t = 0$ . However, there is a problem already with  $\|\nabla \partial_t v(0)\|$ , as has been demonstrated in [41].

### 2.2.1 COMPATIBILITY CONDITIONS AT $t = 0$

To investigate this phenomenon, let us assume that the solution  $\{v, p\}$  is uniformly smooth as  $t \rightarrow 0$ . Then, applying the divergence operator to the Navier-Stokes equations and letting  $t \rightarrow 0$  implies:

(i) in  $\Omega$ :

$$\nabla \cdot (\partial_t v + v \cdot \nabla v) = \nabla \cdot (\nu \Delta v - \nabla p) \quad \rightarrow \quad \nabla \cdot (v^0 \cdot \nabla v^0) = -\Delta p^0,$$

(ii) on  $\partial\Omega$ :

$$\partial_t v + v \cdot \nabla v = \nu \Delta v - \nabla p \quad \rightarrow \quad \partial_t g|_{t=0} + v^0 \cdot \nabla v^0 = \nu \Delta v^0 - \nabla p^0,$$

where  $g$  is the boundary data,  $v^0$  the initial velocity and  $p^0 := \lim_{t \rightarrow 0} p(t)$  the “initial pressure”. Hence, in the limit  $t = 0$ , we obtain an overdetermined Neumann problem for the initial pressure:

$$\Delta p^0 = -\nabla \cdot (v^0 \cdot \nabla v^0) \text{ in } \Omega, \quad (11)$$

$$\nabla p^0|_{\partial\Omega} = \nu \Delta v^0 - \partial_t g|_{t=0} - v^0 \cdot \nabla v^0, \quad (12)$$

including the compatibility condition

$$\partial_\tau p^0|_{\partial\Omega} = \tau \cdot (\nu \Delta v^0 - \partial_t g|_{t=0} - v^0 \cdot \nabla v^0), \quad (13)$$

where  $\tau$  is the tangent direction along  $\partial\Omega$ . If this compatibility is violated, then  $\lim_{t \rightarrow 0} \{\|\nabla^3 v(t)\| + \|\nabla \partial_t v(t)\|\} = \infty$ ; see [41]. We emphasize that (13) together with (11) is a *global* condition which in general cannot be verified for given data. Without (13) being satisfied the maximum degree of regularity is right in the middle of  $\mathbf{H}^2$  and  $\mathbf{H}^3$ ; see [78]. In view of the foregoing discussion, the natural regularity assumption for the nonstationary Navier-Stokes equations (without additional compatibility condition) is

$$v^0 \in \mathbf{J}_1(\Omega) \cap \mathbf{H}^2(\Omega) \quad \Rightarrow \quad \sup_{t \in (0, T]} \{\|\nabla^2 v(t)\| + \|\partial_t v(t)\|\} < \infty. \quad (14)$$

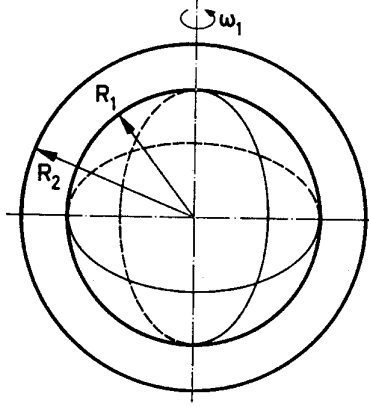
**Example:** *Flow between two concentric spheres (“Taylor problem”)*

Let the inner sphere with radius  $r_{in}$  be accelerated from rest  $v^0 = 0$  with a constant acceleration  $\omega$ , i.e.,  $v|_{\Gamma_{in}} \cdot (n, \tau_\theta, \tau_\phi)^T = r_{in} \cos(\theta)(0, 0, \omega t)^T$  (in polar coordinates), while at the outer sphere, we set  $v|_{\Gamma_{out}} = 0$ . Accordingly, the Neumann problem for the “initial pressure” takes the form

$$\Delta p^0 = 0 \text{ in } \Omega, \quad \partial_n p^0|_{\partial\Omega} = 0,$$

which implies that  $p^0 \equiv \text{const}$ . However, this conflicts with the compatibility condition (13) which in this case reads

$$\partial_\phi p^0|_{\Gamma_{in}} = -\partial_t(v^0 \cdot n_\phi)_{t=0} = -r_{in} \cos(\theta)\omega \neq 0.$$



In order to describe the “natural” regularity of the solution  $\{v(t), p(t)\}$ , as  $t \rightarrow 0$ , and as  $t \rightarrow \infty$ , in [41] a sequence of time-weighted a priori estimates has been proven under the assumption (10) using the weight functions  $\tau(t) = \min(t, 1)$  and  $e^{\alpha t}$ , with fixed  $\alpha > 0$ :

$$\tau(t)^{2n+m-2} \left\{ \|\nabla^m \partial_t^n v(t)\| + \|\nabla^{m-1} \partial_t^n p(t)\| \right\} \leq K, \quad (15)$$

and

$$e^{-\alpha t} \int_0^t e^{\alpha s} \tau(s)^{2n+m-2} \left\{ \|\nabla^m \partial_t^n v(t)\|^2 + \|\nabla^{m-1} \partial_t^n p(t)\|^2 \right\} dt \leq K, \quad (16)$$

for any  $m \geq 2, n \geq 1$ .

**Open problem 2.1:** *Devise a way to construct for any given initial data  $v^0$  (e.g., fitted from experimental data) and any  $\epsilon > 0$  a smooth initial data  $\tilde{v}^0 \in \mathbf{J}_1(\Gamma; \Omega)$ , such that  $\|v^0 - \tilde{v}^0\|_1 \leq \epsilon$ , and the resulting solution of the Navier-Stokes equations satisfies the compatibility condition (13) at  $t = 0$ .*

### 2.3 Outflow boundary conditions

Numerical simulation of flow problems usually requires the truncation of a conceptionally unbounded flow region to a bounded computational domain, thereby introducing artificial boundaries, along which some kind of boundary conditions are needed. The variational formulation (8), (9) does not contain an explicit reference to any “outflow boundary condition”. Suppose that the solution  $v \in v^{in} + \mathbf{H}, p \in L$  is sufficiently smooth. Then, integration by parts on the terms

$$\nu(\nabla v \nabla \phi) - (p, \nabla \cdot \phi) = \int_{\Gamma_{out}} \{\nu \partial_n v - pn\} \phi \, do + (-\nu \Delta v + \nabla p, \phi)$$

yields the already mentioned “natural” condition on the outflow boundary

$$\nu \partial_n v - pn = 0 \text{ on } \Gamma_{out}. \quad (17)$$

This condition has proven to be well suited in modeling (essentially) parallel flows, see, e.g., [46], Turek [93, 97]. It naturally occurs in the variational formulation of the problem if one does not prescribe any boundary condition for the velocity at the outlet suggesting the name “do nothing” boundary condition.

In the following, we present some experiences in choosing the boundary conditions implicitly, through the choice of variational formulations of flow problems used in finite element computations. To fix ideas, let us begin by considering a common test problem, that of calculating nonstationary flow past an obstacle (here an inclined ellipse), situated in a rectangular channel.

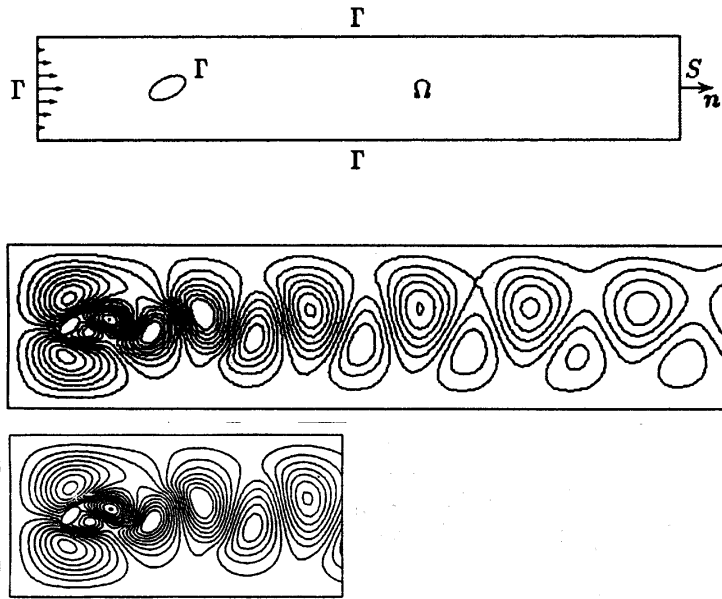


Figure 10: The effect of the “do nothing” outflow boundary condition shown by pressure isolines for unsteady flow around an inclined ellipse at  $Re=500$ ; from [46].

We impose the usual no-slip boundary conditions on the channel walls and on the surface of the ellipse, while a parabolic “Poiseuille” inflow profile is prescribed on the upstream boundary. We denote again by  $\Gamma$  that portion of the boundary on which Dirichlet conditions are imposed. At the downstream boundary  $S = \Gamma_{out}$ , we decide to “do nothing” and leave the solution and the test space free by choosing  $\mathbf{H} = \mathbf{H}_0^1(\Gamma; \Omega)$  and  $L = L^2(\Omega)$ . This results in the free-outflow condition (17). The results of computations based on (17) show a truly remarkable “transparency” of the downstream boundary when it is handled in this way; see Figure 10 where almost no effect of shortening the computational domain is seen.

### 2.3.1 PROBLEMS WITH THE “DO NOTHING” OUTFLOW CONDITION

Although, the “do nothing” outflow boundary condition seems to yield very satisfactory results, one should use it with care. For example, if the flow region contains more than one outlet, like in flows through systems of pipes, undesirable effects may occur, since the “do nothing” condition contains as an additional hidden condition that the mean pressure is zero across the outflow boundary. In fact integrating (17) over any component  $S$  of the outflow boundary (a straight segment) and using the incompressibility constraint  $\nabla \cdot v = 0$  yields

$$\int_S pn \, do = \nu \int_S \partial_n v \, do = -\nu \int_S \partial_t v \, do = v(s_2) - v(s_1) = 0.$$

Here  $s_i$  denote the end points of  $S$  at which  $v(s_i) = 0$ , due to the imposed no-slip condition along  $\Gamma$ . Consequently, the mean pressure over  $S$  must be zero:

$$\int_S p \, do = 0. \quad (18)$$

To illustrate this, let us consider low Reynolds number flow through a junction in a system of pipes, again prescribing a Poiseuille inflow upstream. In Figure 11, we show steady streamlines for computations based on the same variational formulations as above, each with the same inflow, but with varying lengths of pipe beyond the junction. Obviously, making one leg of the pipe longer significantly changes the flow pattern. The explanation of this effect is that by the property (18), in Figure 11 the pressure gradient is greater in the shorter of the two outflow sections, which explains why there is a greater flow through that section.

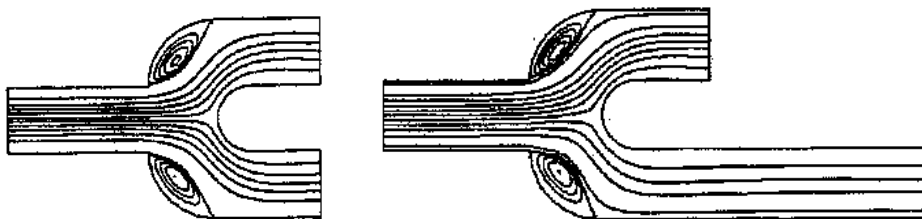


Figure 11: The effect of the “do nothing” outflow boundary condition shown by streamline plots for flow through a bifurcating channel for  $\text{Re} = 20$ ; from [46].

### 2.3.2 MODIFICATION OF TRANSPORT OR DIFFUSION MODEL

The foregoing example suggests that one might consider formulating problems more generally, e.g., in terms of *prescribed pressure drops* or *prescribed fluxes*; we refer to [46] for a thorough development of the corresponding variational formulations. Both choices of boundary conditions lead to well posed formulations of the problem. However, the situation is less satisfactory than in the case of pure Dirichlet boundary conditions. Although the variational problem

looks well set in this situation, surprisingly there is a problem with its well posedness. The related Dirichlet problem of the Navier-Stokes equations, stationary as well as nonstationary, is well known to possess weak solutions (not necessarily unique or stable) for any Reynolds number. The standard argument for this result is based upon the “conservation property”  $(v \cdot \nabla v, v) = 0$  of the nonlinear term, which is obtained by integration by parts and using  $\nabla \cdot v = 0$ . In the case of a “free” boundary this relation is replaced by

$$(v \cdot \nabla v, v) = -\frac{1}{2}(n \cdot v, v^2)_{\Gamma}, \quad \Gamma = \cup_i \Gamma_i, \quad (19)$$

which generally does not allow to bound the energy in the system without a priori knowledge of what is an inflow and what is an outflow boundary. As a consequence, in [46] the existence of a unique solution could be shown, even in two space dimensions, only for sufficiently small data. Kracmar/Neustupa [57] have treated the case of general data by formulating the problem as a variational inequality including the energy bound as a constraint. This still leaves the question open whether one can expect existence of solutions for the original formulation with general data. A positive answer is suggested by numerical tests which do not show any unexpected instability with the discrete analogues of the formulation (8), (9) in the case of higher Reynolds numbers.

One may suspect that this theoretical difficulty can be avoided simply by changing the variational formulation of the problem, i.e., using other variational representations of the transport or diffusion terms. It has been suggested to replace in the momentum equation (8) the Dirichlet form  $(\nabla v, \nabla \phi)$  by  $(D[v], D[\phi])$ , with  $D[v] = \frac{1}{2}(\partial_i v_j + \partial_j v_i)_{i,j=1}^d$  being the deformation tensor. This change has no effect in the case of pure Dirichlet boundary conditions as then the two forms coincide. But in using the “do nothing” approach this modification leads to the outflow boundary condition

$$n \cdot D[v] - pn = 0 \quad \text{on } \Gamma_{out},$$

which may result in a non-physical behavior of the flow. In the case of simple Poiseuille flow the streamlines are bent outward as shown in Figure 12.

Another possible modification is to enforce the conservation property on the transport terms. Using the identity  $\nabla(\frac{1}{2}|v|^2) = v \cdot (\nabla v)^T$ , the transport term can be written in the form

$$v \cdot \nabla v = v \cdot \nabla v - v \cdot (\nabla v)^T + \frac{1}{2} \nabla |v|^2.$$

This leads to a variational formulation in which  $(v \cdot \nabla v, \phi)$  is replaced by

$$\tilde{b}(v, v, \phi) := (v \cdot \nabla v, \phi) - (\phi \cdot \nabla v, v), \quad (20)$$

while the term  $\frac{1}{2}|v|^2$  is absorbed into the pressure. An alternative form is

$$\tilde{b}(v, v, \phi) := \frac{1}{2}(v \cdot \nabla v, \phi) - \frac{1}{2}(v \cdot \nabla \phi, v), \quad (21)$$

which is legitimate because  $\tilde{b}(v, v, \phi) = (v \cdot \nabla v, \phi)$  for  $v \in \mathbf{J}_1(\Omega)$ . Notice that in both cases  $\tilde{b}(w, \phi, \phi) = 0$  for any  $w$ . The corresponding natural outflow boundary conditions are, for (20):

$$\nu \partial_n v - \bar{p} n = 0, \quad (22)$$

with the so-called “Bernoulli pressure”  $\bar{p} = p + \frac{1}{2}|v|^2$ , and for (21):

$$\nu \partial_n v - \frac{1}{2}|n \cdot v|^2 n - p n = 0. \quad (23)$$

Both modifications again result in a non-physical behavior across the outflow boundary; streamlines bent inward as shown in Figure 12. Hence, for physical reasons, it seems to be necessary to stay with the original formulation (8), (9). For a detailed discussion of the boundary conditions, and for an extensive list of references, we refer the reader to Gresho [34] and Gresho and Sani [35].

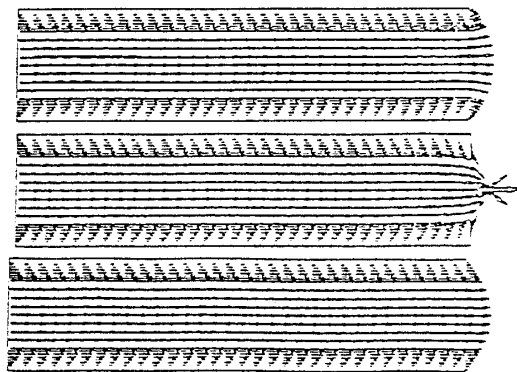


Figure 12: The effect of using the deformation tensor formulation (top) or the symmetrized transport formulations (middle) together with the “do nothing” outflow boundary condition compared to the correct Poiseuille flow (bottom); from [46].

**Open Problem 2.2:** *Prove the existence of global smooth solutions (in 2D) for the original variational Navier-Stokes equations with the “do nothing” outflow boundary condition for general data.*



### 3 Spatial discretization by finite elements

In this section, we recall some basics about the spatial discretization of the incompressible Navier–Stokes equations by the finite element method. The emphasis will be on those types of finite elements which are used in our codes for solving two and three dimensional flow problems, stationary as well as nonstationary. For a general discussion of finite element methods for flow problems, see to Girault/Raviart [29], Pironneau [68], and Gresho/Sani [35].

#### 3.1 Basics of finite element discretization

We begin by a brief introduction to the basics of finite element discretization of elliptic problems, e.g., the Poisson equation in a bounded domain  $\Omega \subset \mathbb{R}^d$  ( $d = 2$  or  $3$ ) with a polyhedral boundary  $\partial\Omega$ ,

$$-\nu\Delta u = f \text{ in } \Omega. \quad (1)$$

We assume homogeneous Dirichlet and Neumann boundary conditions,

$$u|_{\Gamma_D} = 0, \quad \partial_n u|_{\Gamma_N} = 0, \quad (2)$$

along disjoint components  $\Gamma_D$  and  $\Gamma_N$  of  $\partial\Omega$ , where  $\partial\Omega = \overline{\Gamma_D \cup \Gamma_N}$ . The starting point is the variational formulation of this problem in the natural solution space  $H := H_0^1(\Gamma_D; \Omega)$ : Find  $u \in H$  satisfying

$$a(u, \phi) := \nu(\nabla u, \nabla \phi) = (f, \phi) \quad \forall \phi \in H. \quad (3)$$

To discretize this problem, we introduce decompositions, named  $\mathbb{T}_h$ , of  $\bar{\Omega}$  into (closed) cells  $K$  (triangles or quadrilaterals in 2D, and tetrahedra or hexahedra in 3D) such that the usual regularity conditions are satisfied:

- $\bar{\Omega} = \cup\{K \in \mathbb{T}_h\}$ .
- Any two cells  $K, K'$  only intersect in common faces, edges or vertices.
- The decomposition  $\mathbb{T}_h$  matches the decomposition  $\partial\Omega = \Gamma_D \cup \Gamma_N$ .

In the following, we will also allow decompositions with “hanging nodes” in order to ease local mesh refinement. To each of the decompositions  $\mathbb{T}_h$ , there corresponds a mesh-size function  $h = h(x)$  which is piecewise constant such that  $h|_K =: h_K$ . We set  $h_K := \text{diam}(K)$  and denote by  $\rho_K$  the radius of the ball of maximal size contained in  $K$ . We will also use the notation  $h := \max_{K \in \mathbb{T}_h} h_K$ . The family of decompositions  $\{\mathbb{T}_h\}_h$  is said to be (uniformly) “shape regular”, if

$$ch_K \leq \rho_K \leq h_K, \quad (4)$$

and (uniformly) “quasi-uniform”, if

$$\max_{K \in \mathbb{T}_h} h_K \leq c \min_{K \in \mathbb{T}_h} h_K, \quad (5)$$

with some constants  $c$  independent of  $h$ ; see Girault/Raviart [29] or Brenner/Scott [19] for more details of these properties. In the following, we will generally assume *shape-regularity* (unless something else is said). *Quasi-uniformity* is usually not required. Examples of admissible meshes are shown below.

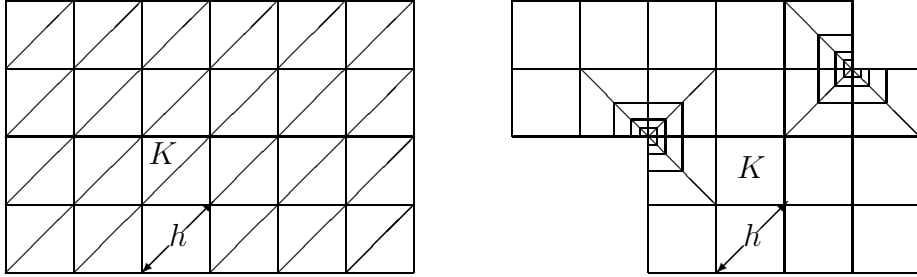


Figure 13: Regular finite element meshes (triangular and quadrilateral)

On the decompositions  $\mathbb{T}_h$ , we consider “finite element spaces”  $H_h \subset H$  defined by

$$H_h := \{v_h \in H, v_h|_K \in P(K), K \in \mathbb{T}_h\},$$

where  $P(K)$  are certain spaces of elementary functions on the cells  $K$ . In the simplest case,  $P(K)$  are polynomial spaces,  $P(K) = P_r(K)$  for some degree  $r \geq 1$ . On general quadrilateral or hexahedral cells, we have to work with “parametric” elements, i.e., the local shape functions are constructed by using transformations  $\psi_K : \hat{K} \rightarrow K$  between the “physical” cell  $K$  and a fixed “reference unit-cell”  $\hat{K}$  by  $v_h|_K(\psi_K(\cdot)) \in P_r(\hat{K})$ . This construction is necessary in general in order to preserve “conformity” (i.e., global continuity) of the cell-wise defined functions  $v_h \in H_h$ . For example, the use of *bilinear* shape functions  $\phi \in \text{span}\{1, x_1, x_2, x_1x_2\}$  on a quadrilateral mesh in 2D employs likewise *bilinear* transformations  $\psi_K : \hat{K} \rightarrow K$ . We will see more examples of concrete finite element spaces below.

In a finite element discretization, “consistency” is expressed in terms of local approximation properties of the shape functions used. For example, in the case of a second-order approximation using *linear* or *d-linear* shape functions, there holds locally on each cell  $K$ :

$$\|v - I_h v\|_K + h_K \|\nabla(v - I_h v)\|_K \leq c_I h_K^2 \|\nabla^2 v\|_K, \quad (6)$$

and on each cell surface  $\partial K$ :

$$\|v - I_h v\|_{\partial K} + h_K \|\partial_n(v - I_h v)\|_{\partial K} \leq c_I h_K^{3/2} \|\nabla^2 v\|_K, \quad (7)$$

where  $I_h v \in H_h$  is the natural “nodal interpolation” of a function  $v \in H \cap H^2(\Omega)$ , i.e.,  $I_h v$  coincides with  $v$  with respect to certain “nodal functionals” (e.g., point values at vertices, mean values on edges or faces, etc.). The “interpolation constant” is usually of size  $c_I \sim 0.1-1$ , depending on the shape of the cell  $K$ .

With the foregoing notation the discrete scheme reads as follows: Find  $u_h \in H_h$  satisfying

$$a(u_h, \phi_h) = (f, \phi_h) \quad \forall \phi_h \in H_h. \quad (8)$$

Combining the two equations (3) and (8) yields the relation

$$a(u - u_h, \phi_h) = 0, \quad \phi_h \in H_h, \quad (9)$$

which means that the error  $e := u - u_h$  is “orthogonal” to the subspace  $H_h$  with respect to the bilinear form  $a(\cdot, \cdot)$ . This essential feature of the finite element Galerkin scheme immediately implies the “best approximation” property

$$\|\nabla e\| = \min_{\phi_h \in H_h} \|\nabla(u - \phi_h)\|. \quad (10)$$

In virtue of the interpolation estimate (6), we obtain the (global) *a priori* convergence estimate

$$\|\nabla e\| \leq c_I h \|\nabla^2 u\| \leq c_I c_S h \|f\|, \quad (11)$$

provided that the solution is sufficiently regular, i.e.,  $u \in H^2(\Omega)$ , satisfying the *a priori* bound

$$\|\nabla^2 v\| \leq c_S \|f\|. \quad (12)$$

In the above model, this is the case if the polygonal domain  $\Omega$  is convex. In case of reduced regularity of  $u$  due to reentrant corners, the order in the estimate is correspondingly reduced. In the case of approximation by higher-order polynomials,  $r \geq 2$ , and higher order of regularity of  $u$ , the estimate (11) shows a correspondingly increased power of  $h$ . The order of  $h$  in the “energy-error” estimate (11) can be improved by shifting to the  $L^2$ -norm. This is done by employing a duality argument (“Aubin-Nitsche trick”); see, e.g., Brenner/Scott [19]. Let  $z \in H$  be the solution of the auxiliary problem

$$-\nu \Delta z = \|e\|^{-1} e \quad \text{in } \Omega, \quad z = 0 \quad \text{on } \partial\Omega, \quad (13)$$

satisfying an *a priori* bound  $\|\nabla^2 z\| \leq c_S$ . Then, there holds

$$\|e\| = (e, -\nu \Delta z) = a(e, z) = a(e, z - I_h z) \leq c_I c_S h \|\nabla e\|, \quad (14)$$

and we conclude the improved *a priori*  $L^2$ -error estimate

$$\|e\| \leq c_I^2 c_S^2 h^2 \|f\|. \quad (15)$$

In order to convert the problems (8) into a form which is amenable to practical computation, we introduce the nodal basis  $\{\phi_h^1, \dots, \phi_h^N\}$ ,  $N = \dim H_h$ , of the space  $H_h$ , defined by  $\phi_h^i(a_j) = \delta_{ij}$ ,  $i, j = 1, \dots, N$ , where  $a_j$  are the nodal points (e.g., the vertices) of the mesh. Then, setting

$$u_h = \sum_{i=1}^N x_i \phi_h^i,$$

problem (8) is equivalent to the linear algebraic system

$$Ax = b, \quad (16)$$

for the “nodal value” vector  $x = (x_i)_{i=1}^N$ . Here, the “stiffness matrix”  $A$  and the “load vector”  $b$  are defined by

$$A := (a(\phi_h^i, \phi_h^j))_{i,j=1}^N, \quad b := ((f, \phi_h^i))_{i=1}^N.$$

In the case of variable coefficients and force the integrals have to be computed by using integration formulas; in our implementations usually Gaussian formulas are used. For the pure diffusion problem the stiffness matrix  $A$  is symmetric and positive definite. Its condition number behaves like  $\kappa(A) = O(h^{-2})$ , where the exponent -2 is determined by the order of the differential operator  $\Delta$  (it is independent of the spatial dimension and the polynomial degree of the finite elements used).

Below, we show a sequence of hexahedral 3D meshes used for computing the “puff-puff flow” mentioned in the Introduction; observe the successively refined approximation of the curved boundary.

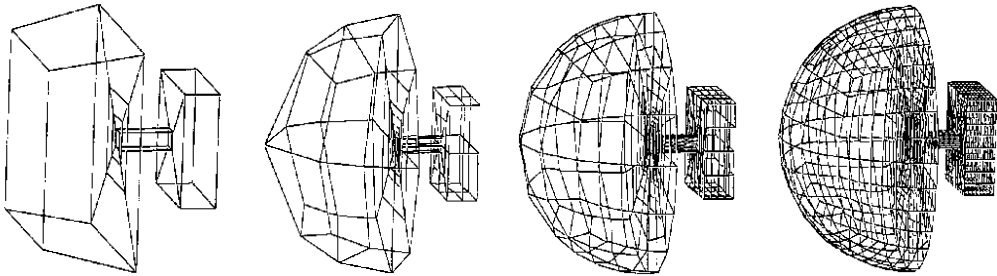


Figure 14: Sequence of successively refined hexahedral meshes for computing the “puff-puff” flow in 3D.

### 3.2 Stokes elements

We consider the *stationary* Navier-Stokes problem as specified in Section 2. In setting up a finite element model of the Navier-Stokes problem, one starts from the variational formulation of the problem: Find  $v \in v^{in} + \mathbf{H}$  and  $p \in L$ , such that

$$a(v, \varphi) + n(v, v, \varphi) + b(p, \varphi) = (f, \varphi) \quad \forall \varphi \in \mathbf{H}, \quad (17)$$

$$b(\chi, v) = 0 \quad \forall \chi \in L. \quad (18)$$

The choice of the function spaces  $\mathbf{H} \subset \mathbf{H}^1(\Omega)$  and  $L \subset L^2(\Omega)$  depends on the specific boundary conditions imposed in the problem to be solved. On a finite element mesh  $\mathbb{T}_h$  on  $\Omega$  with cell width  $h$ , one defines spaces of “discrete” trial and test functions,

$$\mathbf{H}_h \text{ “} \subset \text{” } \mathbf{H}, \quad L_h \subset L.$$

The discrete analogues of (19), (20) then read as follows: Find  $v_h \in v_h^{in} + \mathbf{H}_h$  and  $p_h \in L_h$ , such that

$$a_h(v_h, \varphi_h) + n_h(v_h, v_h, \phi_h) + b_h(p_h, \phi_h) = (f, \varphi_h) \quad \forall \phi_h \in \mathbf{H}_h, \quad (19)$$

$$b_h(\chi_h, v_h) = 0 \quad \forall \chi_h \in L_h, \quad (20)$$

where  $v_h^{in}$  is a suitable approximation of the inflow data  $v^{in}$ . The notation  $\mathbf{H}_h \subset \mathbf{H}$  indicates that in this discretization the spaces  $\mathbf{H}_h$  may be “nonconforming”, i.e., the discrete velocities  $v_h$  are continuous across the interelement boundaries and zero along the rigid boundaries only in an approximate sense; in this case the discrete forms  $a_h(\cdot, \cdot)$ ,  $b_h(\cdot, \cdot)$ ,  $n_h(\cdot, \cdot, \cdot)$  and the discrete “energy norm”  $\|\nabla \cdot\|_h$  are defined in the piecewise sense,

$$\begin{aligned} a_h(\phi, \psi) &:= \sum_{K \in \mathbb{T}_h} \nu(\nabla \phi, \nabla \psi)_K, & b_h(\chi, \phi) &:= \sum_{K \in \mathbb{T}_h} (\chi, \nabla \cdot \phi)_K, \\ n_h(\phi, \psi, \xi) &:= \sum_{K \in \mathbb{T}_h} (\phi \cdot \nabla \psi, \xi)_K, & \|\nabla \phi\|_h &:= \left( \sum_{K \in \mathbb{T}_h} \|\nabla \phi\|_K^2 \right)^{1/2}. \end{aligned}$$

In order that (19), (20) is a stable approximation to (17), (18), as  $h \rightarrow 0$ , it is crucial that the spaces  $\mathbf{H}_h \times L_h$  satisfy a compatibility condition, the so-called “inf-sup” or “Babuska-Brezzi” condition,

$$\inf_{q_h \in L_h} \left\{ \sup_{w_h \in \mathbf{H}_h} \frac{b_h(q_h, w_h)}{\|q_h\| \|\nabla w_h\|_h} \right\} \geq \gamma > 0. \quad (21)$$

Here, the constant  $\gamma$  is required to be independent of  $h$ . This ensures that the problems (19), (20) possess solutions which are uniquely determined in  $\mathbf{H}_h \times L_h$  and stable. Further, for the errors  $e_v := v - v_h$  and  $e_p := p - p_h$ , there hold *a priori* estimates of the form

$$\|\nabla e_v\|_h + \|e_p\| \leq ch \{ \|\nabla^2 v\| + \|\nabla p\| \}. \quad (22)$$

A rigorous convergence analysis of spatial discretization of the Navier-Stokes problem can be found in Girault/Raviart [29] and in [41, 43].

### 3.2.1 EXAMPLES OF STOKES ELEMENTS

Many stable pairs of finite element spaces  $\{\mathbf{H}_h, L_h\}$  have been proposed in the literature (see, e.g., Girault/Raviart [29], Hughes et al. [49] and [77]). Below, two particularly simple examples of quadrilateral elements will be described which have satisfactory approximation properties and are applicable in two as well as in three space dimensions. They can be made robust against mesh degeneration (large aspect ratios) and they admit the application of efficient multigrid solvers. We note that, from the point of view of accuracy, in our context quadrilateral (hexahedral) elements are to be preferred over triangular (tetrahedral) elements because of their superior approximation properties.

Both types of elements may be used in the spatial discretization underlying the discussions in the following sections.

1) *The nonconforming “rotated”  $d$ -linear  $\tilde{Q}_1/P_0$  Stokes element*

The first example is the natural quadrilateral analogue of the well-known triangular nonconforming finite element of Crouzeix/Raviart (see [29]). It was introduced and analyzed in [77] and its two- as well as three-dimensional versions have been implemented in state-of-the-art Navier-Stokes codes (see Turek [93, 96], Schreiber/Turek [83], and Oswald [66]). In two space dimensions, this nonconforming element uses piecewise “rotated” bi-linear (reference) shape functions for the velocities, spanned by  $\{1, x, y, x^2 - y^2\}$ , and piecewise constant pressures. As nodal values one may take the mean values of the velocity vector over the element edges (or, alternatively, its point values at the mid-points of sides) and the mean values of the pressure over the elements. For the precise definition of this element we introduce the set  $\partial\mathbb{T}_h$  of all  $(d-1)$ -faces  $S$  of the elements  $K \in \mathbb{T}_h$ . We set

$$\tilde{Q}_1(K) = \{q \circ \psi_T^{-1} : q \in \text{span}\{1, x_1, x_i^2 - x_{i+1}^2, i = 1, \dots, d\}\}.$$

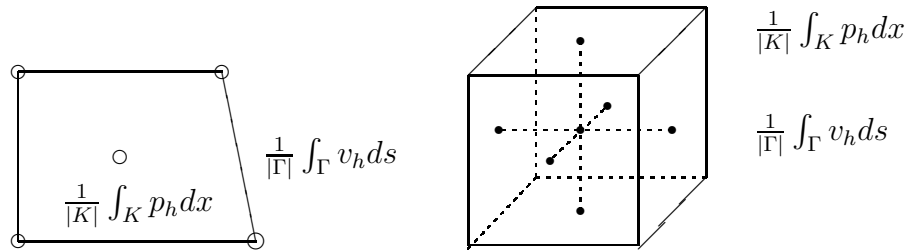
The corresponding finite element spaces are

$$\mathbf{H}_h := \left\{ \begin{array}{l} v_h \in L^2(\Omega)^d : v_{h|K} \in \tilde{Q}_1(K)^d, K \in \mathbb{T}_h, \\ F_S(v_{h|K}) = F_S(v_{h|K'}), S \subset \partial K \cap \partial K', F_S(v_h) = 0, S \subset \Gamma \end{array} \right\},$$

$$L_h := \{q_h \in L : q_{h|K} \in P_0(K), K \in \mathbb{T}_h\},$$

with the nodal functionals

$$F_S(v_h) = |S|^{-1} \int_S v_h \, do, \quad F_K(p_h) = |K|^{-1} \int_K p_h \, dx.$$



Clearly, the spaces  $\mathbf{H}_h$  are non-conforming,  $\mathbf{H}_h \not\subset \mathbf{H}^1(\Omega)^d$ . For the pair  $\{\mathbf{H}_h, L_h\}$  the discrete “inf-sup” stability condition (21) is known to be satisfied on fairly general meshes; see [77] and [13]. For illustration, we recall from [13] the essential steps of the argument.

*Proof of the “inf-sup” stability estimate (21):* Using the continuous “inf-sup” estimate (7), we conclude for an arbitrary  $p_h \in L_h$  that

$$\gamma_0 \|p_h\| \leq \sup_{\phi \in \mathbf{H}} \frac{|b_h(p_h, \phi)|}{\|\nabla \phi\|} \leq \sup_{r_h \phi \in \mathbf{H}_h} \frac{|b_h(p_h, r_h \phi)|}{\|\nabla r_h \phi\|} \sup_{\phi \in \mathbf{H}} \frac{\|\nabla r_h \phi\|}{\|\nabla \phi\|}. \quad (23)$$

where  $r_h\phi \in \mathbf{H}_h$  is an approximation to  $\phi \in \mathbf{H}$  satisfying

$$b_h(\chi_h, \phi - r_h\phi) = 0 \quad \forall \chi_h \in L_h, \quad \|\nabla r_h\phi\| \leq c_* \|\nabla\phi\|. \quad (24)$$

These properties are realized for the  $\tilde{Q}_1/P_0$  Stokes element by the natural nodal interpolation defined by

$$\int_S r_h\phi \, do = \int_S \phi \, do \quad \forall S \in \partial\mathbb{T}.$$

Then, the first relation in (24) is obvious, and the  $\mathbf{H}^1$  stability follows from

$$\begin{aligned} \|\nabla r_h\phi\|_h^2 &= \sum_{K \in \mathbb{T}_h} \left\{ (r_h\phi, \partial_n r_h\phi)_{\partial K} - (r_h\phi, \Delta r_h\phi)_K \right\} \\ &= (\nabla\phi, \nabla r_h\phi)_h + \sum_{K \in \mathbb{T}_h} \left\{ (r_h\phi - \phi, \partial_n r_h\phi)_{\partial K} - (r_h\phi - \phi, \Delta r_h\phi)_K \right\}. \end{aligned}$$

The argument becomes particularly simple for parallelogram meshes. In this case  $\partial_n r_h\phi|_{\Gamma} \equiv \text{const.}$  and  $\Delta r_h\phi|_K \equiv 0$ , such that the last sum vanishes. The general case requires a more involved estimation. Now, the desired “inf-sup” stability estimate follows with the constant  $\gamma = \gamma_0/c_*$ .

As discussed in [77], the stability and approximation properties of the  $\tilde{Q}_1/P_0$  Stokes element depend very sensitively on the degree of deviation of the cells  $K$  from parallelogram shape. Stability and convergence deteriorates with increasing cell aspect ratios. This defect can be cured by using a “non-parametric” version of the element where the reference space  $\tilde{Q}_1(K) := \{q \in \text{span}\{1, \xi, \eta, \xi^2 - \eta^2\}\}$  is defined for each element  $K$  independently with respect to the coordinate system  $(\xi, \eta)$  spanned by the directions connecting the midpoints of sides of  $K$ . This approximation turns out to be robust with respect to the shape of the elements  $K$ , and the convergence estimate (22) remains true. Below, we will relax this requirement even further by allowing the elements to be stretched in one or more (in 3D) directions.

Finally, we mention an important feature of the  $\tilde{Q}_1/P_0$  Stokes element (see [93]): It possesses a “divergence-free” nodal-basis, which allows the elimination of the pressure from the problem resulting in a positive definite algebraic system for the velocity unknowns alone. The reduced algebraic system can be solved by specially adapted multigrid methods; see Turek [93].

## 2) The conforming $d$ -linear $Q_1/Q_1$ Stokes element with pressure stabilization

The second example uses continuous isoparametric  $d$ -linear shape functions for both the velocity and the pressure approximations. The nodal values are just the function values of the velocity and the pressure at the vertices of the mesh, making this approximation particularly attractive in three dimensions. With

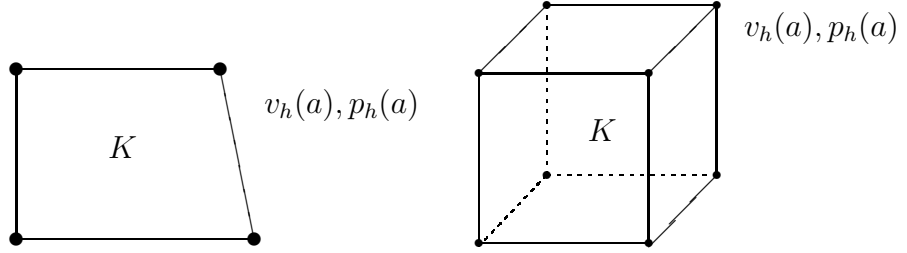
$$\tilde{Q}_1(K) = \{q \circ \psi_T^{-1} : q \in \text{span}\{1, x_i, x_i x_j, i, j = 1, \dots, d\}\},$$

the corresponding finite element spaces are defined by

$$\begin{aligned}\mathbf{H}_h &= \left\{ v_h \in \mathbf{H}_0^1(\Gamma; \Omega)^d : v_{h|K} \in \tilde{Q}_1(K)^d, K \in \mathbb{T}_h \right\}, \\ L_h &= \left\{ q_h \in H^1(\Omega) : q_{h|K} \in \tilde{Q}_1(K), K \in \mathbb{T}_h \right\},\end{aligned}$$

with the nodal functionals ( $a$  vertex of the mesh  $\mathbb{T}_h$ )

$$F_a(v_h) = v_h(a), \quad F_a(p_h) = p_h(a).$$



This combination of spaces, however, would be unstable, i.e., it would violate the condition (21), if used together with the variational formulation (19), (20). In order to get a stable discretization, it was proposed by Hughes et al. [49], to add certain least squares terms in the continuity equation (20) (pressure stabilization method),

$$b(\chi_h, v_h) + c_h(\chi_h, p_h) = g_h(v_h; \chi_h), \quad (25)$$

where

$$\begin{aligned}c_h(\chi_h, p_h) &= \frac{\alpha}{\nu} \sum_{K \in \mathbb{T}_h} h_K^2 (\nabla \chi_h, \nabla p_h)_K, \\ g_h(v_h; \chi_h) &= \frac{\alpha}{\nu} \sum_{K \in \mathbb{T}_h} h_K^2 (\nabla \chi_h, f + \nu \Delta v_h - v_h \cdot \nabla v_h)_K.\end{aligned}$$

The correction terms on the right hand side have the effect that this modification is fully consistent, since the additional terms cancel out if the exact solution  $\{v, p\}$  of problem (17), (18) is inserted. On regular meshes, one obtains a stable and consistent approximation of the Navier-Stokes problem (17), (18), for which a convergence estimate of form (22) holds true. The argument follows a slightly different track than that used above for the nonconforming  $\tilde{Q}_1/P_0$  element; see [13].

*Proof of the “inf-sup” stability estimate (21):* From the continuous stability estimate (7) we conclude that

$$\gamma_0 \|p_h\| \leq \sup_{r_h \phi \in \mathbf{H}_h} \frac{|(p_h, \nabla r_h \phi)|}{\|\nabla r_h \phi\|} \sup_{\phi \in \mathbf{H}} \frac{\|\nabla r_h \phi\|}{\|\nabla \phi\|} + \sup_{\phi \in \mathbf{H}} \frac{|(\nabla p_h, \phi - r_h \phi)|}{\|\nabla \phi\|}, \quad (26)$$



where  $r_h\phi \in \mathbf{H}_h$  is an approximation to  $\phi \in \mathbf{H}$ , satisfying

$$\left( \sum_{K \in \mathbb{T}_h} h_K^{-2} \|\phi - r_h\phi\|_K^2 \right)^{1/2} + \|\nabla r_h\phi\| \leq c_* \|\nabla\phi\|. \quad (27)$$

The existence of such an approximation can be shown by employing an averaged nodal interpolation. From this, we obtain

$$\gamma \|p_h\| \leq c_* \sup_{\phi_h \in \mathbf{H}_h} \frac{|(p_h, \nabla\phi_h)|}{\|\nabla\phi_h\|} + c_* \left( \sum_{K \in \mathbb{T}_h} h_K^2 \|\nabla p_h\|_K^2 \right)^{1/2},$$

which yields the desired stability estimate with the constant  $\gamma = \gamma_0/c_*$ .

It was shown in Hughes et al. [49], and later on in a series of mathematical papers (see, e.g., Brezzi/Pikäranta [21], and the literature cited therein) in the context of a more general analysis of such stabilization methods, that this kind of discretization is numerically stable and of optimal order convergent for many relevant pairs of spaces  $\mathbf{H}_h \times L_h$ .

The stabilized  $Q_1/Q_1$  Stokes element has several important features: With the same number of degrees of freedom it is more accurate than its triangular analogue (and also slightly more accurate than its nonconforming analogue described above). Furthermore, it has a very simple data structure due to the use of the same type of nodal values for velocities and pressure which allows for an efficient vectorization of solution processes. Thanks to the stabilization term in the continuity equation, standard multigrid techniques can be used for solving the algebraic systems with good efficiency (see the discussion in Section 5 below).

We note that the triangular analogue of this element is closely related (indeed almost algebraically equivalent) to the “inf-sup” stable MINI-element (see Brezzi/Fortin [20]) which is based on the standard  $Q_1/Q_1$ -element and stability is achieved by augmenting the velocity space by local cubic bubble functions.

The stabilized  $Q_1/Q_1$ -Stokes element has been implemented in several 2D and 3D Navier-Stokes codes (see, e.g., Harig [38], Becker [7], and Braack [17]). However, it was already reported in Harig [38] that the convergence properties of this element sensitively depend on the parameter  $\alpha$  and may deteriorate on strongly stretched meshes. We will come back to this point below.

### 3.3 The algebraic problems

The discrete Navier-Stokes problem (19), (20), possibly including pressure stabilization (25), has to be converted into an algebraic system which can be solved on a computer. To this end, we choose appropriate local “nodal bases”  $\{\phi_h^i, i = 1, \dots, N_v\}$  of the “velocity space”  $\mathbf{H}_h$ , and  $\{\chi_h^i, i = 1, \dots, N_p\}$  of the

“pressure space”  $L_h$  and expand the unknown solution  $\{v_h, p_h\}$  in the form

$$v_h = v_h^{in} + \sum_{i=1}^{N_v} x_i \phi_h^i, \quad p_h = \sum_{j=1}^{N_p} x_j \chi_h^j.$$

We introduce the following matrices:

$$\begin{aligned} A &= (a_h(\phi_h^i, \phi_h^j))_{i,j=1}^{N_v}, \quad B = (b_h(\chi_h^i, \phi_h^j))_{i,j=1}^{N_p, N_v}, \quad C = (c_h(\chi_h^i, \chi_h^j))_{i,j=1}^{N_p}, \\ N(x) &= (n_h(v_h^{in} + \sum_{k=1}^{N_v} x_k \phi_h^k, \phi_h^i, \phi_h^j) + n_h(\phi_h^i, v_h^{in}, \phi_h^j))_{i,j=1}^{N_v}, \\ b &= ((f, \phi_h^j) - a(v_h^{in}, \phi_h^j) - n_h(v_h^{in}, v_h^{in}, \phi_h^j))_j^{N_v}, \quad c(x) = (g_h(v_h; \chi_h^j))_{j=1}^{N_p}. \end{aligned}$$

Here,  $A$  is the stiffness matrix,  $B$  the “gradient matrix” with the associated “divergence matrix”  $-B^T$ ;  $N(\cdot)$  is the (nonlinear) transport matrix and  $b$  the load vector into which the nonhomogeneous inflow-boundary data have been incorporated. Further,  $C$  is the matrix arising from pressure stabilization and  $c$  the (nonlinear) correction term on the right-hand side. Occasionally, we will use the abbreviation  $A(\cdot) := A + N(\cdot)$ . For later use, we also introduce the velocity and pressure “mass matrices”:

$$M_v = ((\phi_h^i, \phi_h^j))_{i,j=1}^{N_v}, \quad M_p = ((\chi_h^i, \chi_h^j))_{i,j=1}^{N_p}.$$

With this notation the variational problem (19), (20) can equivalently be written in form of an algebraic system for the vectors  $x \in \mathbb{R}^{N_v}$  and  $y \in \mathbb{R}^{N_p}$  of expansion coefficients:

$$Ax + N(x)x + By = b, \tag{28}$$

$$-B^T x + Cy = c(x). \tag{29}$$

Notice that for this system has the structure of a saddle-point problem (for  $C = 0$ ) and is generically nonsymmetric. This poses a series of problems in solving it by iterative methods. This point will be addressed in more detail in Section 5, below.

### 3.4 Anisotropic meshes

In many situations it is necessary to work with (locally) *anisotropic* meshes, i.e., in some areas of the computational domain the cells are stretched in order to better resolve local solution features. Such anisotropies generically occur when tensor-product meshes are used to resolve boundary layers. In this case the mesh  $\mathbb{T}_h$  is no longer “quasi-regular” and the discretization may strongly depend on the deterioration of the cells measured in terms of “cell aspect ratios”. On such meshes three different phenomena occur:

- The constant  $c_I$  in the interpolation estimates (6), (7) may blow up.
- The constant  $\gamma$  in the “inf-sup” stability estimate (21) may become small.
- The conditioning of the algebraic system may deteriorate.

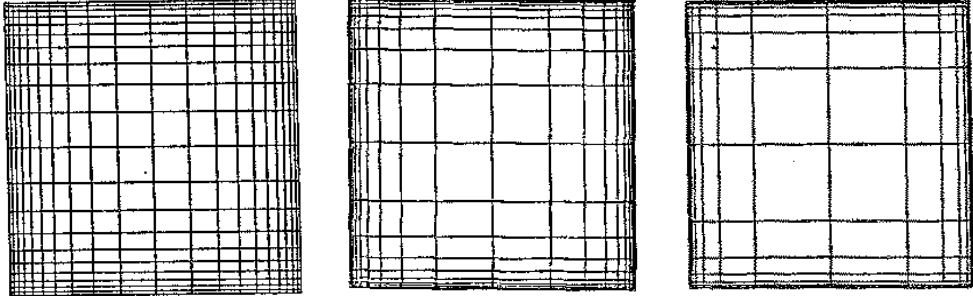
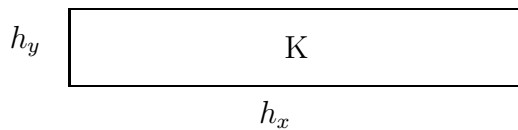


Figure 15: A sequence of locally anisotropic tensor-product meshes with aspect ratios  $\sigma_h = 10, 100, 1000$ , for computing the driven-cavity flow.

It is known that for most of the lower-order finite elements (including the elements considered here) the local interpolation estimates remain valid even on highly stretched elements (“maximum angle condition” versus “minimum angle condition”). Accordingly, the failure of the considered Stokes elements on stretched meshes is not so much a problem of consistency but rather one of stability. Hence, we will discuss the stability in more detail; the approximation aspects have been systematically analyzed in Apel/Dobrowolski [2], Apel [3], and the literature cited therein. The main technical difficulty arises from the deterioration of the “inverse inequality” for finite elements  $\|\nabla\phi_h\|_K \leq ch_K^{-1}\|\phi_h\|_K$  on stretched cells. Further, the solution of the resulting algebraic systems, e.g., by multigrid methods, becomes increasingly difficult. For simplicity, we concentrate the following discussion on the special case of cartesian tensor-product cells as shown in the figures above and below; here the “cell aspect ratio” is defined by  $\sigma_K = h_x/h_y$  and the maximum “mesh aspect ratio” by  $\sigma_h := \max_{K \in \mathbb{T}_h} \sigma_K$ . We consider aspect ratios of size  $\sigma_h \approx 1 - 10^4$ .



As a model problem, we consider the stationary Stokes equations

$$-\nu\Delta u + \nabla p = f, \quad \nabla \cdot u = 0, \quad \text{in } \Omega, \quad (30)$$

with homogeneous Dirichlet boundary conditions  $u|_{\partial\Omega} = 0$ , on a bounded polygonal domain  $\Omega \in \mathbb{R}^2$ . Using the notation introduced above, the finite element formulation of this problem reads as follows:

$$a_h(u_h, \phi_h) + b_h(p_h, \phi_h) = (f, v) \quad \forall \phi_h \in \mathbf{H}_h, \quad (31)$$

$$b_h(u_h, \chi_h) = 0 \quad \forall \chi_h \in L_h. \quad (32)$$

1. First, we consider the *nonconforming*  $\tilde{Q}_1/P_0$  Stokes approximation which uses “rotated” bilinear shape functions for the velocities and piecewise constants for the pressure. Above, we have introduced its “non-parametric” version where local cell coordinates  $\{\xi_K, \eta_K\}$  are used for defining the local shape functions on each cell as  $v_{h|K} \in \text{span}\{1, \xi_K, \eta_K, \xi_K^2 - \eta_K^2\}$ . In this way one obtains a discretization which is robust with respect to deviations of the cell from parallelogram shape. It has been shown in [12] that this non-parametric element can be modified to be also robust with respect to increasing aspect ratio. This modification employs a scaling of the local coordinate system according to the cell aspect ratio  $\sigma_K := h_x/h_y$ ,

$$v_{h|K} \in \text{span}\{1, \xi_K, \sigma_K \eta_K, \xi_K^2 - \sigma_K^2 \eta_K^2\}.$$

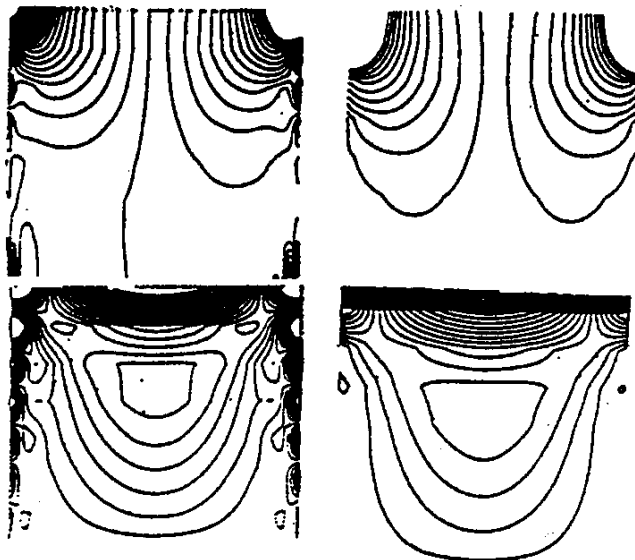


Figure 16: Pressure and velocity norm isolines for a driven-cavity computation with the standard isotropic  $\tilde{Q}_1/P_0$  element (left) compared to the anisotropically scaled version (right); from Becker [7].

Furthermore, the “inf-sup” stability estimate (21) is preserved on such meshes with a constant  $\gamma$  independent of the mesh aspect ratio  $\sigma_h$ . To demonstrate that this scaling is actually necessary for the stability of this element, we show in Figure 16 the results of a “driven cavity” calculation on meshes with  $\sigma_h = 32$  using the standard isotropic approximation compared to the anisotropically scaled version. The instability caused by the large aspect ratio exhibits spurious pressure peaks and vortices along the boundary.

2. Next, we consider the stabilized  $Q_1/Q_1$ -Stokes approximation which uses continuous (isoparametric) bilinear shape functions for both the velocity and

the pressure. As seen before, this discretization becomes “inf-sup” stable if the discrete model is augmented by a least-squares term of the form

$$(\nabla \cdot u_h, \chi_h) + c(p_h, \chi_h) = \text{“correction terms”}. \quad (33)$$

On quasi-uniform meshes, we obtain a stable and consistent approximation of the Stokes problem but this approximation sensitively depends on the choice of the form  $c(\cdot, \cdot)$  and may deteriorate on strongly stretched meshes. Again, the approximation property of the  $Q_1/Q_1$  element is not the problem. The interpolation estimates (6) and (7) remain valid also on high-aspect-ratio meshes (as defined above) with constants independent of  $\sigma_h$ . However, the proper design of the stabilization (33) is delicate. We consider the following three different choices for the stabilizing bilinear form:

$$c(p, q) = \begin{cases} c_1(p, q) = \alpha \sum_{K \in \mathbb{T}_h} |K| (\nabla p, \nabla q)_K, \\ c_2(p, q) = \alpha \sum_{K \in \mathbb{T}_h} h_K^2 (\nabla p, \nabla q)_K, \\ c_3(p, q) = \alpha \sum_{K \in \mathbb{T}_h} \{h_x^2 (\partial_x p, \partial_x q)_K + h_y^2 (\partial_y p, \partial_y q)_K\}. \end{cases}$$

The form  $c_1(\cdot, \cdot)$  is built in analogy to the MINI–element, since condensation of the bubble functions leads directly to the cell-wise scaling factor  $|K|$ . We see that  $c_1(\cdot, \cdot)$  gets smaller with increasing  $\sigma_h$ , an undesirable effect which is avoided by  $c_2(\cdot, \cdot)$ . Finally,  $c_3(\cdot, \cdot)$  distinguishes between the different coordinate directions which requires the use of a local coordinate system in the definition of the stabilization. By a local “inverse estimate” for bilinear functions on (arbitrary) rectangles we get the stability relation  $c_3(p_h, p_h) \leq \|p_h\|^2$ , which appears necessary for the stabilization to achieve uniformity with respect to the mesh aspect ratio. This may be seen by writing the discrete system (31), (33) in matrix notation

$$\begin{bmatrix} A & B \\ -B^T & C_i \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} b \\ c \end{bmatrix},$$

where  $C_i$  corresponds to the stabilizing bilinear form  $c_i(\cdot, \cdot)$ . The Schur complement of the main diagonal block  $A$  is  $\Sigma = C_i - B^T A^{-1} B$ . Then, the stability constant  $\gamma$  in (33) is given by (see [13]):

$$\gamma^2 = \lambda_{\min}(M^{-1}\Sigma), \quad (34)$$

where  $M$  denotes the mass matrix of the pressure space  $L_h$  (piecewise constants in this case). This correspondence can be used in order to experimentally determine the dependence of the stability constant  $\gamma$  on the various parameters of the discretization, particularly the cell aspect ratio. We may detect  $\gamma$  by counting the number of cg–iterations (preconditioned by the mass matrix  $M$ ) needed to invert the Schur complement. The convergence rate  $\rho$  of the cg–iteration applied to the preconditioned Schur complement  $M^{-1}\Sigma$  is linked to the condition number  $\kappa = \text{cond}(M^{-1}\Sigma)$  by the following well-known formula

$$\rho \approx 2 \left( \frac{1 - 1/\sqrt{\kappa}}{1 + 1/\sqrt{\kappa}} \right).$$

These test calculations use a sequence of anisotropic grids obtained by one-directional refinements. The results are given in Table 1.

Table 1: Number of cg-iterations; from Becker [7].

$\sigma$	2	4	8	16	32	64	128
$c_1$	8	18	39	98	559	*	*
$c_2$	8	18	39	88	193	*	*
$c_3$	8	16	29	31	29	27	24

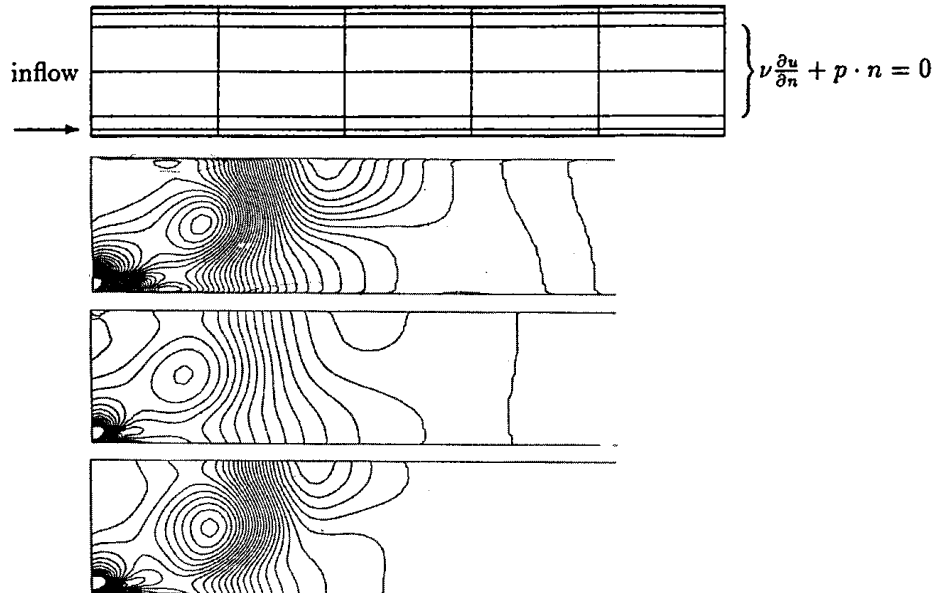


Figure 17: Pressure isolines for a jet flow in a channel calculation with the  $Q_1/Q_1$  element using isotropic stabilization (top and middle) compared to the anisotropic stabilization (bottom); from Becker [7].

The interpretation of these observations is as follows: The increase of the stability constant for  $c_1(\cdot, \cdot)$  stems from the fact, that  $\gamma \approx \sigma^{-2} \rightarrow 0$  with increasing aspect ratio, whereas the bad behavior of  $c_2(\cdot, \cdot)$  can be explained by the growth of  $\lambda_{max}(\Sigma) \approx \sigma^2$  due to fact that we only have  $c_2(p, q) \leq \sigma \|p\| \|q\|$ . We also see that the anisotropic stabilization by  $c_3(\cdot, \cdot)$  leads to an aspect-ratio-independent behavior. Similar effects are also observed for the accuracy of the different stabilizations; see Becker [7] and [13].

**Open Problem 3.1:** *Prove the approximation property (27) on general meshes with arbitrary aspect ratio  $\sigma_h$ . The special case of tensor-product meshes has been treated in Becker [7] (see also Apel [3]).*

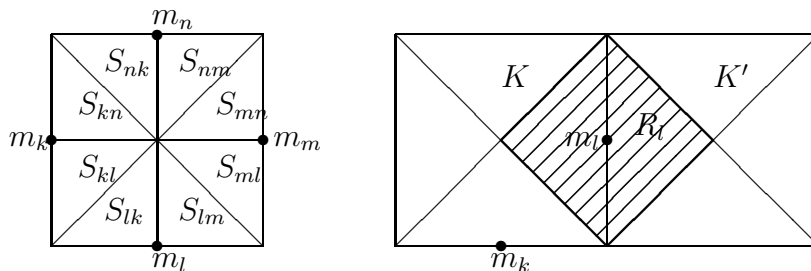
### 3.5 Treatment of dominant transport

In the case of higher Reynolds numbers (e.g.,  $\text{Re} > 1000$  for the 2-D driven cavity, and  $\text{Re} > 100$  for the flow around an cylinder) the finite element models (19), (20) or (19), (25) may become unstable since they essentially use central-differences-like discretization of the advective term. This “instability” most frequently occurs in form of a drastic slow-down or even break-down of the iteration processes for solving the algebraic problems; in the extreme case the possibly existing “mathematical” solution contains strongly oscillatory components without any physical meaning. In order to avoid these effects some additional numerical damping is required. The use of simple first-order artificial viscosity is not advisable since it introduces too much numerical damping. Below, we describe two approaches used in the context of finite element discretization: i) an adaptive upwinding, and ii) the streamline diffusion method. Alternative techniques are the “characteristics Galerkin method” (for nonstationary flows) and the “discontinuous finite element method” which require major changes in the discretization and will therefore not be discussed here; for references see Pironneau [67], Morton [63] and Johnson [51].

#### 3.5.1 UPWINDING

In the finite element context “upwinding” can be defined in a quite natural way; see, e.g., Tobiska/Schieweck [90] and Turek [97], and the literature cited therein. Here, the upwinding effect is accomplished in the evaluating of the advection term through shifting integration points into the upwind direction. This modification leads to system matrices which have certain M-matrix properties and are therefore amenable to efficient and robust solution techniques. This is widely exploited in the finite element codes described in Schreiber/Turek [83], Turek [97] and Schieweck [82].

Following [97], we briefly describe the upwind strategy for the nonconforming “rotated” bilinear Stokes element. Each quadrilateral  $K \in \mathbb{T}_h$  is divided into eight barycentric fragments  $S_{ij}$ , and for each edge  $\Gamma_l$  and mid point  $m_l$  on  $\Gamma_l$  the “lumping region”  $R_l$  is defined by  $R_l := \cup_{k \in \Lambda_l} S_{lk}$ , where  $\Lambda_l = \{k, m_l \text{ and } m_k \text{ belong to the same element } K\}$ . The boundary of the lumping region  $R_l$  consists of the edges  $\Gamma_{lk} := \partial S_{lk} \cap \partial S_{kl}$ , i.e.,  $\partial R_l = \cup_{k \in \Lambda_l} \Gamma_{lk}$ . In this way we obtain an edge oriented partition of the mesh domain  $\Omega_h = \cup_l R_l$ .



A modification of the nonlinear form  $n(u_h, v_h, w_h)$  is now defined by

$$\tilde{n}_h(u_h, v_h, w_h) := \sum_{K,l} (1 - \lambda_{lk}(u_h)) (v_h(m_k) - v_h(m_l)) w_h(m_l) \int_{\Gamma_l} u_h \cdot n_{lk} \, ds,$$

where  $\lambda_{lk}$  are parameters depending on the local flux direction. Setting

$$x := \frac{1}{\nu} \int_{\Gamma_{lk}} u_h \cdot n_{lk} \, ds,$$

possible choices are

$$\begin{aligned} \lambda_{lk} &:= \left\{ \begin{array}{l} 1, \text{ for } x \geq 0 \\ 0, \text{ for } x < 0 \end{array} \right\} \quad (\text{“simple upwinding”}), \\ \lambda_{lk} &:= \left\{ \begin{array}{l} (1/2 + x)/(1 + x), \text{ for } x \geq 0 \\ 1/(2 - 2x), \text{ for } x < 0 \end{array} \right\} \quad (\text{“Samarskij upwinding”}), \end{aligned}$$

It can be shown (see Tobiska/Schieweck [90]) that this upwind scheme is of first order accurate and, what is most important, the main diagonal blocks of the corresponding system matrix  $A + \tilde{N}(\cdot)$  become M-matrices. This is the key property for its inversion by fast multigrid algorithms.

The described upwind discretization can generically be extended to the three dimensional case. An analogous construction is possible for the conforming  $Q_1/Q_1$  element with pressure stabilization also in two as well as in three dimensions; see Harig [38].

### 3.5.2 STREAMLINE DIFFUSION

The idea of “streamline diffusion” is to introduce artificial diffusion acting only in the transport direction while maintaining the second-order consistency of the scheme. This can be achieved in various ways, by augmenting the test space by direction-oriented terms resulting in a “Petrov-Galerkin method”, or by adding certain least-squares terms to the discretization. For the (stationary) Navier-Stokes problem, we propose the following variant written in terms of pairs  $\{\phi, \chi\} \in \mathbf{H} \times L$ : Find  $v_h \in v_h^{in} + \mathbf{H}_h$  and  $p_h \in L_h$ , such that

$$\begin{aligned} a_h(v_h, \phi_h) + n_h(v_h, v_h, \phi_h) + b_h(p_h, \nabla \cdot \phi_h) + s_h(\{v_h, p_h\}, \{\phi_h, \chi_h\}) \\ = (f, \phi_h) + r_h(\{v_h, p_h\}, \{\phi_h, \chi_h\}) \end{aligned} \quad (35)$$

for all  $\{\phi_h, \chi_h\} \in \mathbf{H} \times L_h$ , where, with some “reference velocity”  $\bar{v}_h$ ,

$$\begin{aligned} s_h(\{v_h, p_h\}, \{\phi_h, \chi_h\}) &= \sum_{K \in \mathbb{T}_h} \delta_K \left\{ (\nabla p_h + v_h \cdot \nabla v_h, \nabla \chi_h + \bar{v}_h \cdot \nabla \phi_h)_K \right. \\ &\quad \left. + (\nabla \cdot v_h, \nabla \cdot \phi_h)_K \right\}, \\ r_h(\{v_h, p_h\}, \{\phi_h, \chi_h\}) &= \sum_{K \in \mathbb{T}_h} \delta_K (f + \nu \Delta v_h, \nabla \chi_h + \bar{v}_h \cdot \nabla \phi_h)_K. \end{aligned}$$



The stabilization parameters  $\delta_K$  are chosen according to

$$\delta_K = \min \left\{ \frac{h_K^2}{\nu}, \frac{h_K}{|\bar{v}|_K} \right\}. \quad (36)$$

This discretization contains several features. The first term in the sum

$$\sum_{K \in \mathbb{T}_h} \delta_K \{ (\nabla p_h, \nabla \chi_h)_K + (v_h \cdot \nabla v_h, \bar{v}_h \cdot \nabla \phi_h)_K + (\nabla \cdot v_h, \nabla \cdot \phi_h)_K \}$$

stabilizes the pressure-velocity coupling for the conforming  $Q_1/Q_1$  Stokes element, the second term stabilizes the transport operator, and the third term enhances mass conservation. The other terms introduced in the stabilization are correction terms which guarantee second-order accuracy for the stabilized scheme. Theoretical analysis shows that this kind of Galerkin stabilization actually leads to an improvement over the standard upwinding scheme, namely an error behavior like  $\mathcal{O}(h^{3/2})$  for the finite elements described above; see Tobiska/Verfürth [91], and also Braack [17] where the same kind of stabilization has been applied for weakly compressible flows with chemical reactions. For linear convection-diffusion problems the streamline diffusion method is known to have even  $\mathcal{O}(h^2)$  convergence on fairly general meshes; see [107].

**Open Problem 3.2:** *Derive a strategy for choosing the stabilization parameter  $\delta_K$  in the streamline diffusion method on general meshes with arbitrarily large aspect ratio  $\sigma_h$ .*

**Open Problem 3.3:** *The streamline diffusion method (like the least-squares pressure stabilization) leads to a scheme which lacks local mass conservation. Recently, for convection-diffusion problems an alternative approach has been proposed which uses a “discontinuous” Galerkin approximation on the transport term and combines (higher-order) upwinding features with local mass conservation. The extension of this method to the incompressible Navier-Stokes equations (and its practical realization) has yet to be done.*

## 4 Time discretization and linearization

We now consider the *nonstationary* Navier-Stokes problem: Find  $v \in v^{in} + \mathbf{H}$  and  $p \in L$ , such that  $v(0) = v^0$  and

$$(\partial_t v, \varphi) + a(v, \varphi) + n(v, v, \varphi) + b(p, \varphi) = (f, \varphi), \quad \forall \varphi \in \mathbf{H}, \quad (1)$$

$$b(\chi, v) = 0, \quad \forall \chi \in L. \quad (2)$$

The choice of the function spaces  $\mathbf{H} \subset \mathbf{H}^1(\Omega)^d$  and  $L \subset L^2(\Omega)$  depends again on the specific boundary conditions chosen for the problem to be solved; see the discussion in Section 2.

In the past, explicit time stepping schemes have been commonly used in nonstationary flow calculations, mainly for simulating the transition to steady state limits. Because of the severe stability problems inherent to this approach (for moderately sized Reynolds numbers) the very small time steps required prohibited the accurate solution of really time dependent flows. In implicit time stepping one distinguishes traditionally between two different approaches called the “Method of Lines” and the “Rothe Method”.

### 4.1 The Rothe Method

In the “Rothe Method”, at first, the time variable is discretized by one of the common time differencing schemes; for a general account of such schemes see, e.g., Thomée [89]. For example, the backward Euler scheme leads to a sequence of Navier-Stokes-type problems of the form:

$$k_n^{-1}(v^n - v^{n-1}, \phi) + a(v^n, \phi) + n(v^n, v^n, \phi) + b(p^n, \phi) = (f^n, \phi), \quad (3)$$

$$b(\chi, v^n) = 0. \quad (4)$$

for all  $\{\phi, \chi\} \in \mathbf{H} \times L$ , where  $k_n = t_n - t_{n-1}$  is the time step. Each of these problems is then solved by some spatial discretization method as described in the preceding section. This provides the flexibility to vary the spatial discretization, i.e. the mesh or the type of trial functions in the finite element method, during the time stepping process. In the classical Rothe method the time discretization scheme is kept fixed and only the size of the time step may change. The question of how to deal with varying spatial discretization within a time-stepping process while maintaining higher-order accuracy and conservation properties is currently subject of intensive research. It is essential to do the mesh-transfer by  $L^2$  projection which is costly, particularly in 3D if full remeshing is used in each time step, but is easily manageable if only meshes from a family of hierarchically ordered meshes are used.

### 4.2 The Method of Lines

The traditional approach to solving time-dependent problems is the “Method of Lines”. At first, the spatial variable is discretized, e.g. by a finite element

method as described in the preceding section leading to a system of ordinary differential equations of the form:

$$M\dot{x}(t) + Ax(t) + N(x(t))x(t) + By(t) = b(t), \quad (5)$$

$$-B^T x(t) + Cy(t) = c(t), \quad t > 0, \quad (6)$$

with the initial value  $x(0) = x^0$ . The mass matrix  $M$ , the stiffness matrix  $A$  and the gradient matrix  $B$  are as defined above in Section 3. The matrix  $C$  and the right-hand side  $c$  stem from the pressure stabilization when using the conforming  $Q_1/Q_1$  Stokes element. Further, the (nonlinear) matrix  $N(\cdot)$  is thought to contain also all terms arising through the transport stabilization by upwinding or streamline diffusion. For abbreviation, we will sometimes use the notation  $A(\cdot) := A + N(\cdot)$ .

For solving this ODE system, one now applies a time differencing scheme. The most frequently used schemes are the so-called ‘‘One-Step- $\theta$  Schemes’’:

*One-step  $\theta$ -scheme:* Step  $t_{n-1} \rightarrow t_n$  ( $k =$  time step):

$$\begin{aligned} [M + \theta k A^n]x^n + B y^n &= [M - (1-\theta)k A^{n-1}]x^{n-1} + \theta k b^n + (1-\theta)k b^{n-1} \\ -B^T x^n + C y^n &= c^n, \end{aligned}$$

where  $x^n \approx x(t_n)$  and  $A^n := A(x^n)$ . Special cases are the ‘‘forward Euler scheme’’ for  $\theta = 0$  (first-order explicit), the backward Euler scheme for  $\theta = 1$  (first-order implicit, strongly A-stable), and the most popular Crank-Nicolson scheme for  $\theta = 1/2$  (second-order implicit, A-stable). These properties can be seen by applying the method to the scalar model equation  $\dot{x} = \lambda x$ . In this context it is related to a rational approximation of the exponential function of the form

$$R_\theta(-\lambda) = \frac{1 - (\theta - \frac{1}{2})\lambda}{1 + \theta\lambda} = e^{-\lambda} + \mathcal{O}((\theta - \frac{1}{2})|\lambda|^2 + |\lambda|^3), \quad |\lambda| \leq 1.$$

The most robust implicit Euler scheme ( $\theta = 1$ ) is very dissipative and therefore not suitable for computing really nonstationary flow. In contrast, the Crank-Nicolson scheme has only very little dissipation but occasionally suffers from unexpected instabilities caused by the possible occurrence of rough perturbations in the data which are not damped out due to the only weak stability properties of this scheme (not *strongly* A-stable). This defect can in principle be cured by an adaptive step size selection but this may enforce the use of an unreasonably small time step, thereby increasing the computational costs. For a detailed discussion of this issue see [71]. A good time-stepping scheme of the described type should possess the following properties:

- *A-stability* ( $\Rightarrow$  local convergence):  $|R(-\lambda)| \leq 1$
- *Global stability* ( $\Rightarrow$  global convergence):

$$\overline{\lim}_{Re \lambda \rightarrow \infty} |R(-\lambda)| \leq 1 - \mathcal{O}(k).$$

- *Strong A-stability* ( $\Rightarrow$  smoothing property):

$$\overline{\lim}_{Re \lambda \rightarrow \infty} |R(-\lambda)| \leq 1 - \delta < 0.$$

- *Low dissipation* ( $\Rightarrow$  energy preservation):

$$|R(-\lambda)| = 1 - \mathcal{O}(|Im \lambda|), \text{ for } Re \lambda \rightarrow 0.$$

Alternative schemes of higher order are based on the (diagonally) implicit Runge-Kutta formulas or the backward differencing multi-step formulas, both being well known from the ODE literature. These schemes, however, have not yet found wide applications in flow computations, mainly because of their higher complexity and storage requirements compared with the Crank-Nicolson scheme. Also less theoretical analysis is available for these methods when applied to large stiff systems. Some comparison of their stability and approximation properties is made in [72]; see also [65]. However, there is still another method which is an attractive alternative to the Crank-Nicolson method, the so-called “Fractional-Step- $\theta$  Scheme” originally proposed by Glowinski [30] and Bristeau et al. [22].

*Fractional-Step- $\theta$ -scheme:* (three substeps:  $t_{n-1} \rightarrow t_{n-1+\theta} \rightarrow t_{n-\theta} \rightarrow t_n$ )

$$\begin{aligned} (1) \quad & [M + \alpha\theta k A^{n-1+\theta}]x^{n-1+\theta} + \theta k B y^{n-1+\theta} = [M - \beta\theta k A^{n-1}]x^{n-1} + \theta k b^{n-1}, \\ & -B^T x^{n-1+\theta} + C y^{n-1+\theta} = c^{n-1+\theta}, \\ (2) \quad & [M + \beta\theta' k A^{n-\theta}]x^{n-\theta} + \theta' k B y^{n-\theta} = [M - \alpha\theta' k A^{n-1+\theta}]x^{n-1+\theta} + \theta' k b^{n-\theta}, \\ & -B^T x^{n-\theta} + C y^{n-\theta} = c^{n-\theta}, \\ (3) \quad & [M + \alpha\theta k A^n]x^n + \theta k B y^n = [M - \beta\theta k A^{n-\theta}]x^{n-\theta} + \theta k b_h^{n-\theta}, \\ & -B^T x^n + C y^n = c^n. \end{aligned}$$

In the ODE context this scheme reduces to a rational approximation of the exponential function of the form

$$R_\theta(-\lambda) = \frac{(1 - \alpha\theta'\lambda)(1 - \beta\theta\lambda)^2}{(1 + \alpha\theta\lambda)^2(1 + \beta\theta'\lambda)} = e^{-\lambda} + \mathcal{O}(|\lambda|^3), \quad |\lambda| \leq 1.$$

Here  $\theta = 1 - \sqrt{2}/2 = 0.292893\dots$ ,  $\theta' = 1 - 2\theta$ ,  $\alpha \in (1/2, 1]$ , and  $\beta = 1 - \alpha$ , in order to ensure second-order accuracy, and strong A-stability,

$$\overline{\lim}_{Re \lambda \rightarrow \infty} |R_\theta(-\lambda)| = \frac{\beta}{\alpha} < 1.$$

For the special choice  $\alpha = (1 - 2\theta)/(1 - \theta) = 0.585786\dots$ , there holds  $\alpha\theta = \beta\theta'$  which is useful in building the system matrices in the three substeps. This scheme was first proposed in form of an operator splitting scheme separating the two complications “nonlinearity” and “incompressibility” within each cycle  $t_n \rightarrow t_{n+\theta} \rightarrow t_{n+1}$ . However, the Fractional-Step- $\theta$  scheme has also very

attractive features as a pure time-stepping method. It is strongly A-stable, for any choice of  $\alpha \in (1/2, 1]$ , and therefore possesses the full smoothing property in the case of rough initial data, in contrast to the Crank–Nicolson scheme (case  $\alpha = 1/2$ ). Furthermore, its amplification factor has modulus  $|R(-\lambda)| \approx 1$ , for  $\lambda$  approaching the imaginary axis (e.g.,  $|R(-0.8i)| = 0.9998\dots$ ), which is desirable in computing oscillatory solutions without damping out the amplitude. Finally, it also possesses very good approximation properties, i.e., one cycle of length  $(2\theta + \theta')k = k$  provides the same accuracy as three steps of the Crank–Nicolson scheme with total step length  $k/3$ ; for more details on this comparison see [72] and [65].

We mention some theoretical results on the convergence of these schemes. For the Crank–Nicolson Scheme combined with spatial discretization as described in Section 3, an optimal-order convergence estimate

$$\|v_h^n - v(\cdot, t_n)\| = \mathcal{O}(h^2 + k^2) \quad (7)$$

has been given in [44]. This estimate requires some additional stabilization of the scheme but then holds under realistic assumptions on the data of the problem. A similar result has been shown by Müller [64] for the Fractional-Step  $\theta$ -Scheme. Due to its stronger damping properties (*strong* A-stability) this scheme does not require extra stabilization.

#### 4.2.1 COMPUTATIONAL TESTS

Below, we present some results of the computational comparison between the backward Euler scheme, the Crank–Nicolson scheme and the Fractional-Step- $\theta$  scheme. The flow configuration is shown in Figure 18: flow around an inclined plate in the cross-section of a channel at  $\text{Re} = 500$ . The spatial discretization is by the nonconforming “rotated” bilinear Stokes element described in Section 3 on a uniformly refined mesh with 13,000 cells.

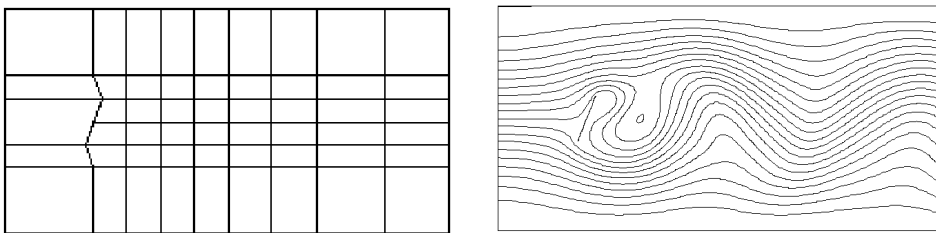


Figure 18: Configuration of plate-flow test, coarse mesh and streamline plot.

The first test concerns accuracy. Figure 19 shows that the backward Euler (BE) scheme is not suitable for computing time-periodic flows with acceptable time-step widths, while the Crank–Nicolson (CN) and the Fractional-Step- $\theta$  (FS) scheme show equally satisfactory results. This similar accuracy is further confirmed by comparing a more sensitive error quantity (mean pressure) in

Figure 20. Finally, we look at the stability of the schemes. Figure 21 demonstrates the lack of robustness of the Crank-Nicolson scheme combined with linear time-extrapolation in the nonlinearity for larger time steps.

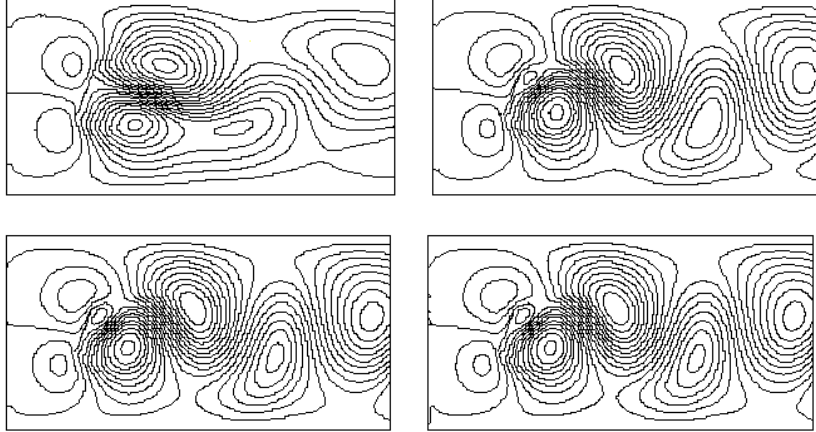


Figure 19: Pressure isolines of the plate-flow test: BE scheme with  $3k = 1$  (top left), BE scheme with  $3k = 0.1$  (top right), CN scheme with  $3k = 1$  (bottom left), FS scheme with  $k = 1$  (bottom right); from [65].

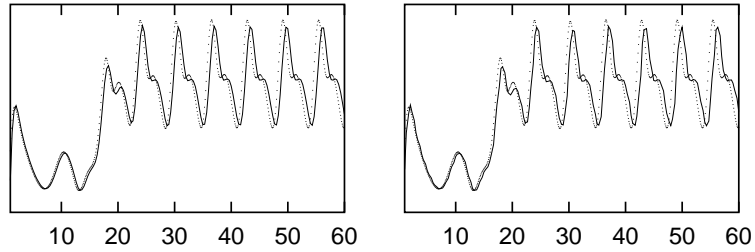


Figure 20: Mean pressure plots for the plate test with fully implicit treatment of the nonlinearity; left: CN scheme with  $3k = 0.33$ ; right: FS scheme with  $k = 0.33$ , both compared to a reference solution (dotted line); from [65].

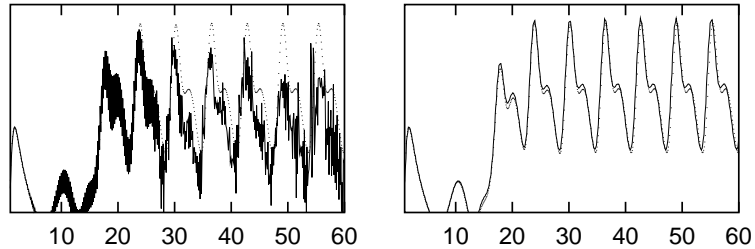


Figure 21: Mean pressure plots for the plate test with linear time-extrapolation; left: CN scheme with  $3k = 0.11$ ; right: FS scheme with  $k = 0.11$ , both compared to a reference solution (dotted line); [65].

### 4.3 Splitting and projection schemes

As already mentioned, the Fractional-Step- $\theta$  scheme was originally introduced as an operator splitting scheme in order to separate the two main difficulties in solving problem (1) namely the nonlinearity causing nonsymmetry and the incompressibility constraint causing indefiniteness. At that time, handling both complications simultaneously was not feasible. Therefore, the use of operator splitting seemed the only way to compute nonstationary flows. Using the notation from above, the splitting scheme reads as follows (suppressing here the terms stemming from pressure stabilization).

*Splitting-Fractional-Step- $\theta$  Scheme:*

$$\begin{aligned}
 (1) \quad & [M + \alpha\theta kA]x^{n-1+\theta} + \theta kBy^{n-1+\theta} = [M - \beta\theta kA]x^{n-1} + \theta kb^{n-1} - \\
 & \quad \quad \quad - \theta kN^{n-1}x^{n-1}, \\
 & B^T x^{n-1+\theta} = 0, \\
 (2) \quad & [M + \beta\theta' kA^{n-\theta}]x^{n-\theta} = [M - \alpha\theta' kA^{n-1+\theta}]x^{n-1+\theta} - \theta' kBy^{n-\theta} + \theta' kb^{n-\theta}, \\
 & \quad \quad \quad \dots \\
 (3) \quad & [M + \alpha\theta kA]x^n + \theta kBy^n = [M - \beta\theta kA]x^{n-\theta} - \theta kN^{n-\theta}x^{n-\theta} + \theta kb^{n-\theta}, \\
 & B^T x^n = 0.
 \end{aligned}$$

The first and last step solve *linear* Stokes problems treating the nonlinearity explicitly, while in the middle step a nonlinear Burgers-type problem (without incompressibility constraint) is solved. The symmetric form of this scheme follows the ideas from Strang [87], in order to achieve a second-order splitting approximation. The results of Müller [64] suggest that the optimal-order convergence estimate (7) remains true also for this splitting scheme. However, a complete proof under realistic assumptions is still missing.

**Open Problem 4.1:** *Prove that the Splitting-Fractional-Step- $\theta$  scheme is actually second order accurate for all choices of the parameter  $\alpha \in (\frac{1}{2}, 1]$ .*

In these days, the efficient solution of the *nonlinear* incompressible Navier-Stokes equations is standard by the use of new multigrid techniques. Hence, the splitting of nonlinearity and incompressibility is no longer an important issue. One of these new approaches uses the Fractional-Step- $\theta$  scheme in combination with the idea of a “projection method” due to Chorin [24]; for a survey see Gresho/Sani [35]. Finally, Turek [95] (see also [97]) has designed the “Discrete Projection Fractional-Step- $\theta$  scheme” as component in his solver for the nonstationary Navier-Stokes problem.

Next, we address the problem of how to deal with the incompressibility constraint  $\nabla \cdot v = 0$ . The traditional approach is to decouple the continuity equation from the momentum equation through an iterative process (again “operator splitting”). There are various schemes of this kind in the literature referred to, e.g., as “quasi-compressibility method”, “projection method”, “SIMPLE method”, etc. All these methods are based on the same principle

idea. The continuity equation  $\nabla \cdot \mathbf{v} = 0$  is supplemented by certain stabilizing terms involving the pressure, e.g.,

$$\nabla \cdot v + \varepsilon p = 0, \quad (8)$$

$$\nabla \cdot v - \varepsilon \Delta p = 0, \quad \partial_n p|_{\partial\Omega} = 0, \quad (9)$$

$$\nabla \cdot v + \varepsilon \partial_t p = 0, \quad p|_{t=0} = 0, \quad (10)$$

$$\nabla \cdot v - \varepsilon \partial_t \Delta p = 0, \quad \partial_n p|_{\partial\Omega} = 0, \quad p|_{t=0} = 0, \quad (11)$$

where the small parameter  $\varepsilon$  is usually taken as  $\varepsilon \approx h^\alpha$ , or  $\varepsilon \approx k^\beta$ , depending on the purpose of the procedure. For example, (8) corresponds to the classical “penalty method”, and (9) is the simplest form of the “least squares pressure stabilization” scheme (11) described above, with  $\varepsilon \approx h^2$  in both cases. Further, (10) corresponds to the “quasi-compressibility method”.

These approaches are closely related to the classical “projection method” of Chorin [24]. Since this method used to be particularly attractive for computing nonstationary incompressible flow, we will discuss it in a some detail. For simplicity consider the case of homogeneous Dirichlet boundary conditions,  $v|_{\partial\Omega} = 0$ . The projection method reads as follows. For an admissible initial value  $v^0$ , choose a time step  $k$ , and solve for  $n \geq 1$ :

(i)  $\tilde{v}^n \in \mathbf{H}$  (implicit “Burgers step”):

$$k^{-1}(\tilde{v}^n - v^{n-1}) - \nu \Delta \tilde{v}^n + \tilde{v}^n \cdot \nabla \tilde{v}^n = f^n. \quad (12)$$

(ii)  $v^n = P\tilde{v}^n \in \mathbf{J}_0(\Omega)$  (“Projection step”):

$$\nabla \cdot v^n = 0, \quad n \cdot v^n|_{\partial\Omega} = 0. \quad (13)$$

Here, the function space  $\mathbf{J}_0(\Omega)$  is obtained through the completion of the space  $\{\phi \in \mathcal{D}(\Omega), \nabla \cdot \phi \equiv 0\}$  of solenoidal test functions with respect to the norm of  $\mathbf{L}^2(\Omega)$ . This time stepping scheme can be combined with any spatial discretization method, e.g. the finite element methods described in Section 3. The projection step (ii) can equivalently be expressed in the form

$$(ii') \quad v^n = \tilde{v}^n + k \nabla \tilde{p}^n, \quad (14)$$

with some “pressure”  $\tilde{p}^n \in H^1(\Omega)$ , which is determined by the properties

$$(ii'') \quad \Delta \tilde{p}^n = k^{-1} \nabla \cdot \tilde{v}^n, \quad \partial_n \tilde{p}^n|_{\partial\Omega} = 0. \quad (15)$$

This amounts to a Poisson equation for  $\tilde{p}^n$  with zero Neumann boundary conditions. It is this non-physical boundary condition,  $\partial_n \tilde{p}^n|_{\partial\Omega} = 0$ , which has caused a lot of controversial discussion about the value of the projection method. Nevertheless, the method has proven to work well for representing the velocity field in many flow problems of physical interest (see, e.g. Gresho [32] and Gresho/Chan [33]). It is very economical as it requires in each time step



only the solution of a (nonlinear) advection-diffusion system for  $v^n$  (of Burgers equation type) and a scalar Neumann problem for  $\tilde{p}^n$ . Still, it was argued that the pressure  $\tilde{p}^n$  were a mere fictitious quantity without any physical relevance. It remained the question: How can such a method work at all? A challenging problem for mathematical analysis!

The first convergence results for the projection method was already given by Chorin, but concerned only cases with absent rigid boundaries (all-space or spatially periodic problems). Later on, qualitative convergence was shown even for the pressure, but in a measure theoretical sense, too weak for practical purposes. Only recently, stronger results on the error behavior of this method have been obtained (see, e.g., Shen [84, 85] as well as [73], and the literature cited therein). The best known error estimate is

$$\|v^n - v(t_n)\|_\Omega + \|\tilde{p}^n - p(t_n)\|_{\Omega'} = \mathcal{O}(k), \quad (16)$$

where  $\Omega' \subset\subset \Omega$  is a subdomain with positive distance to the boundary  $\partial\Omega$ . This shows that the quantities  $\tilde{p}^n$  are indeed reasonable approximations to the pressure  $p(t_n)$ , and finally confirms that Chorin's original method is a first-order time stepping scheme for the incompressible Navier-Stokes problem. The key to this result is the re-interpretation of the projection method in the context of the "pressure stabilization methods". To this end, one inserts the quantity  $v^{n-1} = \tilde{v}^{n-1} - k\nabla\tilde{p}^{n-1}$  into the momentum equation, obtaining

$$k^{-1}(\tilde{v}^n - \tilde{v}^{n-1}) - \nu\Delta\tilde{v}^n + (\tilde{v}^n \cdot \nabla)\tilde{v}^n + \nabla\tilde{p}^{n-1} = f^n, \quad \tilde{v}^n|_{\partial\Omega} = 0, \quad (17)$$

$$\nabla \cdot \tilde{v}^n - k\Delta\tilde{p}^n = 0, \quad \partial_n \tilde{p}^n|_{\partial\Omega} = 0. \quad (18)$$

This looks like an approximation of the Navier-Stokes equations involving a first-order (in time) "pressure stabilization" term, i.e., the projection method can be viewed as a pressure stabilization method with a global stabilization parameter  $\varepsilon = k$ , and an explicit treatment of the pressure term. Moreover, it appears that the pressure error is actually confined to a small boundary strip of width  $\delta \approx \sqrt{\nu k}$  and decays exponentially into the interior of  $\Omega$ . In fact, it was conjectured that, setting  $d(x) = \text{dist}(x, \partial\Omega)$ ,

$$|\tilde{p}^n(x) - p(x, t_n)| \leq c \exp\left(-\alpha \frac{d(x)}{\sqrt{\nu k}}\right) \sqrt{k} + \mathcal{O}(k). \quad (19)$$

This conjecture is supported by numerical experiments for the pressure stabilization method applied to the stationary Stokes problem and by some model situation analysis in E/Liu [25]. The analysis of this boundary layer phenomenon requires the study of the singularly perturbed Neumann problem

$$(\nu^{-1}\nabla \cdot \Delta_D^{-1}\nabla - \varepsilon\Delta_N)q = \varepsilon\Delta p, \quad \text{in } \Omega, \quad \partial_n q|_{\partial\Omega} = \partial_n p|_{\partial\Omega}, \quad (20)$$

where  $\Delta_D$  denotes the Laplacian operator corresponding to Dirichlet boundary conditions. Clearly,  $\nabla \cdot \Delta_D^{-1}\nabla$  is a zero-order operator mapping  $L^2(\Omega)$

into  $L^2(\Omega)$ . For this problem, one would like to know a decay estimate of the form

$$\|q\|_{\Omega_\delta} \leq c \exp\left(-\alpha \frac{\delta}{\sqrt{\epsilon}}\right) \|\nabla p\| + c\nu\epsilon \|\Delta p\|, \quad (21)$$

for interior subdomains  $\Omega_\delta := \{x \in \Omega, \text{dist}(x, \partial\Omega) > \delta\}$ , from which a point-wise result like (19) could be inferred. Such an estimate could be proven in [73] only for the case that the “global” operator  $\nabla \cdot \Delta_D^{-1} \nabla$  is replaced by the “local” identity operator. In the general case the corresponding result is still an open problem.

**Open Problem 4.2:** *Prove an analogue of the a priori decay estimate (21) for the non-local operator  $\nabla \cdot \Delta_D^{-1} \nabla$ .*

The occurrence of the pressure boundary layer is demonstrated in Figure 22 for a simple model problem on the unit square with known polynomial solution. It is even possible to recover the optimal-order accuracy of the pressure,  $\mathcal{O}(h^2)$ , at the boundary by postprocessing, e.g. by linear or quadratic extrapolation of pressure values from the interior of the domain; see Figure 23 and Blum [15] for more details on this matter.

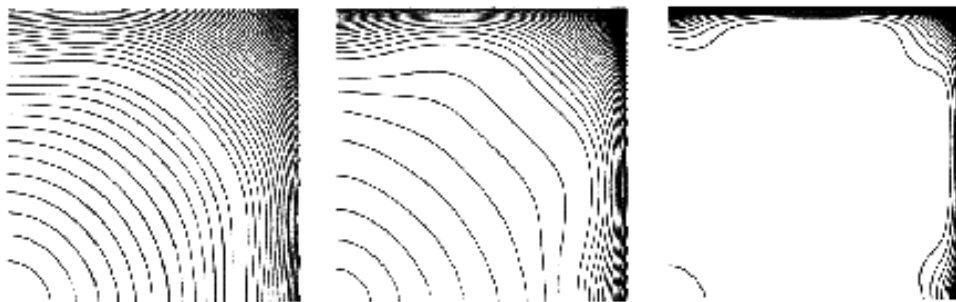


Figure 22: Sequence of pressure-error isolines obtained by the Chorin scheme with  $k = 2.5 \cdot 10^{-2}$ ,  $6.25 \cdot 10^{-3}$ ,  $1.56 \cdot 10^{-3}$  (model problem with  $\nu = 1$  on the square); from Prohl [69].

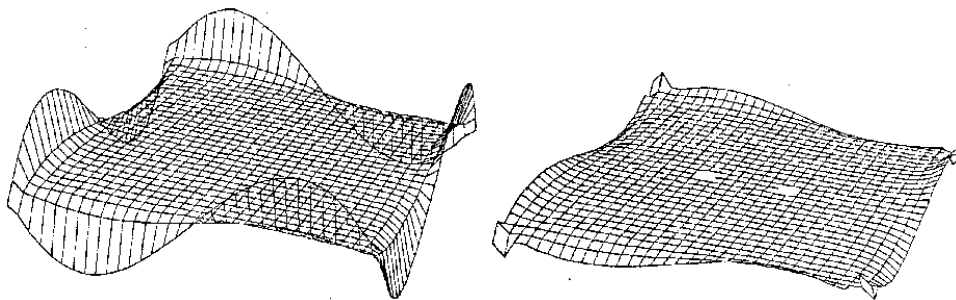


Figure 23: Pressure error plots for a polynomial Stokes solution before (left) and after (right) correction by extrapolation to the boundary; from Blum [15].

An important step towards the solution of the “boundary layer problem” has been made in Prohl [69, 70] by introducing the “Chorin-Uzawa scheme”, which reads as follows:

(i) Implicit “Burgers step”:

$$k^{-1}(\tilde{v}^n - v^{n-1}) - \Delta \tilde{v}^n + \tilde{v}^n \cdot \nabla \tilde{v}^n + \nabla(\tilde{p}^{n-1} - p^{n-1}) = f^n, \quad \tilde{v}^n|_{\partial\Omega} = 0.$$

(ii) Pressure Poisson problem:

$$\Delta \tilde{p}^n = k^{-1} \nabla \cdot \tilde{v}^n, \quad \partial_n \tilde{p}^n|_{\partial\Omega} = 0.$$

(iii) Pressure and velocity update:

$$v^n = \tilde{v}^n - k \nabla \tilde{p}^n, \quad p^n = p^{n-1} - \alpha \nabla \cdot \tilde{v}^n, \quad \alpha < 1.$$

The reference to the name “Uzawa” is due to the fact that this scheme partially resembles the structure of the well-known Uzawa algorithm for solving stationary saddle-point problems; see Girault/Raviart [29]. It corresponds to a quasi-compressibility method using the regularization

$$\nabla \cdot \tilde{v}^n + \alpha^{-1} k \partial_t p^n = 0. \quad (22)$$

This splitting scheme does not introduce a singular perturbation in the pressure equation and is therefore supposed to be free of any spatial boundary layer. However, it suffers from a “boundary layer” at time  $t = 0$  in case of natural initial data not satisfying unrealistic global compatibility conditions; recall Section 2 for a discussion of such conditions. The conjectured suppression of the spatial pressure boundary layer by the Chorin-Uzawa scheme is confirmed by computational tests; see the example presented in Figure 24. A supporting analysis has been given in Prohl [70] for a modification of the Chorin-Uzawa method to a “multi-component scheme” which allows for the convergence estimate

$$\|p^n - p(t_n)\| \leq ck, \quad t_n \geq 1. \quad (23)$$

Figures 24 show pressure error plots obtained for a given polynomial solution on the unit-square with viscosity  $\nu = 1$ ; the time step is  $k = 1/100$  and the spatial discretization uses the  $Q_1/Q_1$  Stokes element with pressure stabilization on a uniform mesh with mesh-size  $h = 1/64$ .

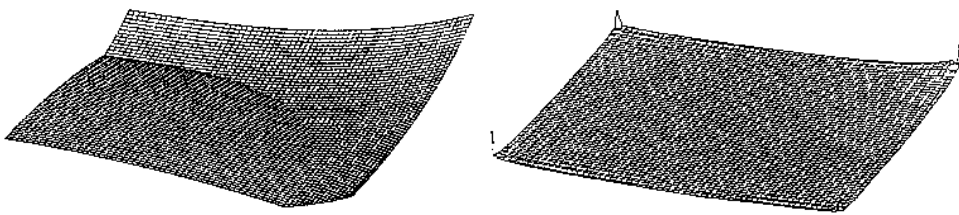


Figure 24: Pressure error plots for a polynomial solution produced by the standard Chorin scheme (left) and the Chorin-Uzawa scheme (right); from Prohl [69, 70].

The projection approach can be extended to formally higher order projection methods. The most popular example is Van Kan's Method [100]: For admissible starting values  $v^0$  and  $p^0$  compute, for  $n \geq 1$  and some  $\alpha \geq \frac{1}{2}$ :

(i)  $\tilde{v}^n \in \mathbf{H}$  (second order implicit Burgers step), satisfying

$$k^{-1}(\tilde{v}^n - v^{n-1}) - \frac{1}{2}\nu\Delta(\tilde{v}^n + v^{n-1}) + \tilde{v}^n \cdot \nabla \tilde{v}^n + \nabla p^{n-1} = f^{n-1/2};$$

(ii)  $p^n \in H^1(\Omega)$ :  $v^n = \tilde{v}^n - \alpha k \nabla(p^n - p^{n-1})$ .

A careful examination of this scheme shows that it can also be interpreted as a certain pressure stabilization method using a stabilization of the form

$$\nabla \cdot v - \alpha k^2 \partial_t \Delta p = 0, \quad \text{in } \Omega, \quad \partial_n p|_{\partial\Omega} = 0, \quad (24)$$

i.e., this method may be viewed as an (implicit) quasi-compressibility method of the form (11) with  $\varepsilon \approx k^2$ ; see [74] and Shen [86].

The projection method may be combined with any of the spatial discretizations described in Section 3. It should be remarked that the simple first-order Chorin scheme is not suitable for computing stationary limits since it has not the form of a fixed-point iteration. In contrast to that, the second-order scheme of Van Kan is designed as a defect-correction iteration and may therefore lead to convergence towards steady state solutions. However, in this case it requires extra pressure stabilization when used together with the conforming  $Q_1/Q_1$  Stokes element; in fact the stabilizing effect of the projection step disappears as  $\alpha k^2 \partial_t \Delta p \rightarrow 0$ .

**Open Problem 4.3:** *The efficient use of projection methods requires an automatic time-step-size control which should monitor deviation from the fully coupled solution. Design such a method for high-order schemes.*

## 5 Solution of the algebraic systems

In this section, we describe solution algorithms for the finite-dimensional problems arising from the discretization presented in the previous sections. These problems form huge and highly nonlinear algebraic systems with a characteristic structure which is exploited by the algorithms. The solution procedure consists of several nested loops. Usually the outermost loop is an implicit time-iteration. In each time step, the arising nonlinear system is solved by a quasi-Newton or defect-correction iteration. The discretization by finite elements leads to a sparse structure in the system matrices which is exploited by the iterative solution method. Even in the case of the Laplace operator (which is always a part of the system), the inversion by a direct solver or a simple iterative scheme like the “conjugate gradient” (CG) method is prohibitive due to the bad conditioning of the matrix with decreasing mesh size. Therefore, the use of multigrid methods is mandatory, either directly as solvers or as preconditioners for a robust iterative schemes like the “generalized minimal residual” (GMRES) algorithm. Since the systems to be solved are in general non-symmetric and indefinite, the construction of “good” multigrid algorithms requires special care.

### 5.1 Linearization

The time stepping schemes described above require in each time step the solution of nonlinear systems of the form

$$[\sigma M + \nu A + N(v)]v + Bp = g, \quad (1)$$

$$-B^T v + \epsilon C p = c, \quad (2)$$

where  $\sigma = (\theta k)^{-1}$  and (on a quasi-uniform mesh)  $\epsilon \sim h^2$ . The operators involved correspond to differential operators as follows:

$$\begin{aligned} M &\sim id., & A &\sim -\text{diag}(\Delta_D), & N(v) &\sim v \cdot \nabla, \\ B &\sim \nabla, & -B^T &\sim \div, & C &\sim -\Delta_N, \end{aligned}$$

where  $\Delta_D$  and  $\Delta_N$  denote the Laplacian operator combined with (homogeneous) Dirichlet or Neumann boundary conditions, respectively. The right-hand sides  $g$  and  $c$  contain information from the preceding time level. Here and below, the same notation is used for the (discrete) velocity  $v$  and pressure  $p$  and the corresponding nodal vectors. The following iteration schemes are formulated on the continuous level without incorporating stabilization, i.e., we set  $\epsilon = 0$  and  $c = 0$ .

*a) Newton method:*

Starting from some initial values  $v^0, p^0 \in \mathbf{H} \times L$  (for example, taken from the preceding time level), one iterates:

1. Defect:  $d^l = g - (\sigma M + \nu A + N(v^l))v^l - Bp^l$ .

2. Correction:  $[\sigma M + \nu A + N'(v^l)]w^l + Bq^l = d^l, \quad B^T w^l = 0.$
3. Update:  $v^{l+1} = v^l + \lambda_l w^l, \quad p^{l+1} = p^l + \lambda_l q^l \quad (\lambda_l \text{ damping factor}).$

This iteration has been observed to converge very fast, provided that it converges at all. The problem is its lack of robustness particularly in the case of larger Reynolds numbers. This is due to the structure of the operator to be inverted in each iteration step:

$$N'(v)w = v \cdot \nabla w + w \cdot \nabla v.$$

It contains a reaction term  $w \cdot \nabla v$  which effects the main diagonal of the system matrix in an uncontrolled manner and may cause divergence of the iteration. This problem may be avoided by simply dropping the reaction term in the Jacobian which results in the following fixed-point defect correction iteration.

*b) Fixed-point defect correction:*

Starting from some initial values  $v^0, p^0 \in \mathbf{H} \times L$  (taken again from the preceding time level), one iterates:

1. Defect:  $d^l = g - (\sigma M + \nu A + N(v^l))v^l - Bp^l.$
2. Correction:  $[\sigma M + \nu A + N(v^l)]w^l + Bq^l = d^l, \quad B^T w^l = 0.$
3. Update:  $v^{l+1} = v^l + \lambda_l w^l, \quad p^{l+1} = p^l + \lambda_l q^l \quad (\lambda_l \text{ damping factor}).$

In this scheme the preconditioning operator  $\tilde{A}'(v^l) = v^l \cdot \nabla$  only contains a transport term which can be stabilized by any of the methods described above: upwinding, streamline diffusion, etc. Normally, within the time stepping scheme, only a few (usually 3-5) steps of the defect correction iteration are necessary for reducing the initial residual down to the level of the discretization error. This is our method of choice used in the codes mentioned in the Introduction.

*c) Nonlinear multigrid iteration:*

The multigrid method can be applied directly to the nonlinear system; see Hackbusch [36]. This may lead to faster convergence but its optimization is difficult and depends very much on the particular problem. Because of this lack of robustness, we do not advocate “nonlinear” multigrid for solving the Navier-Stokes equations.

*d) Nonlinear least-squares cg method:*

The (nonlinear) *least squares cg-method* for solving systems like (1) has been proposed by Glowinski/Periaux [31]. Starting from an initial guess  $x^0$ , a sequence of approximate solutions  $(x^l)_{l \geq 0}$  is obtained by minimizing the least squares functional

$$\|\nabla w\|^2 \rightarrow \min! \tag{3}$$

where  $w$  is determined by  $\{v, p\}$  through the equation

$$[\sigma M + \nu A]w - Bq = \text{“defect” of } \{v, p\}, \quad B^T w = 0. \quad (4)$$

It can be seen that each nonlinear cg-step actually requires only the solution of three linear Stokes problems which can be efficiently done by linear multigrid techniques. This method is very robust as it is based on the minimization of a positive functional, but the speed of convergence drastically slows down for larger Reynolds numbers. For example, the 3-D driven cavity problem can be solved by the stationary version of the least squares cg method up to about  $\text{Re} = 2000$ ; for further details, see [38].

## 5.2 Solution of the linearized problems

The problem to be solved has the form

$$\begin{bmatrix} S & B \\ -B^T & \epsilon C \end{bmatrix} \begin{bmatrix} v \\ p \end{bmatrix} = \begin{bmatrix} g \\ c \end{bmatrix}, \quad \mathcal{A} = \begin{bmatrix} S & B \\ -B^T & \epsilon C \end{bmatrix}, \quad (5)$$

where, with some initial guess  $\bar{v}$ ,

$$S = \sigma M + \nu A + N(\bar{v}).$$

The difficulty with this system is that the matrix  $\mathcal{A}$  is neither symmetric nor definite. It is usually too large for the application of direct solvers (like the LU decomposition by Gaussian elimination) and also the traditional iterative methods (like SOR iteration or Krylov space schemes) do not work sufficiently well. This suggests the use of multigrid methods which are particularly suited on very fine meshes. However, the construction of efficient multigrid algorithms for solving the indefinite system (5) is not at all straightforward. Therefore, as a simpler alternative the Schur complement approach has become popular which will be described in the following subsection.

### 5.2.1 SCHUR-COMPLEMENT ITERATION

In the system matrix  $\mathcal{A}$  the main block  $S$  is regular and usually robust to be inverted. Hence, the velocity unknowns may be eliminated from the system by inverting  $S$  which leads to:

$$[B^T S^{-1} B + \epsilon C]p = B^T S^{-1} g + c, \quad v = S^{-1}(g - Bp). \quad (6)$$

The “Schur-complement” matrix  $\Sigma = B^T S^{-1} B + \epsilon C$  is regular. Neglecting the influence of the nonlinear term  $N(\bar{v})$ , its condition number behaves like

$$\text{cond}(\Sigma) = \mathcal{O}(h^{-2}) \text{ for } \nu k \ll h^2, \quad \text{cond}(\Sigma) = \mathcal{O}(1) \text{ for } \nu k \gg h^2.$$

This suggests the use of iterative methods for its inversion, e.g., Krylov space methods like the GMRES or the bi-cg-stab method. In the essentially non-stationary case,  $\nu k \geq h^2$ , only a few iteration steps suffice. In nonstationary

computations where  $\nu k \ll h^2$ , preconditioning by an approximation of the Neumann-type operator  $B^T M^{-1} B$  is necessary. In each iteration step the operator  $S^{-1}$  has to be evaluated which amounts to solving a linear transport-diffusion problem. In the case of the nonconforming  $\tilde{Q}_1/P_0$  Stokes elements combined with upwind stabilization of advection  $S$  becomes an M-matrix. This facilitates the iterative inversion of  $\Sigma$ , particularly by multigrid methods. The non-exact inversion of  $\Sigma$  makes the step  $q \rightarrow \tilde{\Sigma}^{-1} q$  a preconditioning step within the iteration for inverting  $\Sigma$ . Hence, the number of inner iteration steps should be kept fixed during the whole solution process. Another strategy for compensating for the error in the evaluation of  $S^{-1}$  is to embed the outer iteration (6) into a defect correction process; see Bank, et al., [6]. The convergence usually deteriorates for increasing Reynolds number, because of loss of “symmetry”, and for decreasing time step  $k$ , because of the bad conditioning of the operator  $B^T B \sim \Delta$ . For larger Reynolds number the convergence of the Schur complement iteration becomes slow and special preconditioning is necessary. The construction of effective preconditioners is not easy since the operator  $\Sigma$  is not available as an explicit matrix. Another stability problem occurs on meshes containing cells with large aspect ratio. Because of this lack of robustness, the Schur complement method has less potential than the direct multigrid approach which will be described below.

**Open Problem 5.1:** *Derive a formula for the dependence of the conditioning of the Schur complement operator  $\Sigma = B^T S^{-1} B + \epsilon C$  on the Reynolds number and on the mesh aspect ratio  $\sigma_h$ .*

### 5.3 Linear multigrid solution

The main idea underlying a multigrid method is the fast reduction of high-frequency error components by “cheap” relaxation (“smoothing”) on the fine meshes and the reduction of the remaining low-frequency error components by defect correction on coarser meshes (“coarse-grid correction”); see Hackbusch [36] and Wesseling [104], for an introduction to multigrid methods.

#### 5.3.1 MULTIGRID AS A PRECONDITIONER

Let  $\mathcal{A}$  be the finite element system matrix of the linearized equation (5) or an appropriate approximation. While the theory of multigrid is well developed for scalar elliptic equations, the situation is less clear for complicated systems as considered in this paper. From mathematical analysis, we know that the use of the multigrid method as a preconditioner in an outer iteration (e.g., a Krylov space method such as GMRES) requires less restrictive assumptions than using the multigrid method directly as a solver. In the first case, denoting by  $\mathcal{M}$  the action of a multigrid step, it is sufficient to have an upper bound for the condition of the product  $\mathcal{M}\mathcal{A}$ , whereas in the second case, the eigenvalues of the iteration matrix  $\mathcal{B} = I - \mathcal{M}\mathcal{A}$  have to be uniformly bounded away from one. Therefore, we choose the first option to construct



a robust iteration scheme for the system (5). As basic solver, one may use the “generalized minimal residual method” (GMRES) for the preconditioned matrix  $\mathcal{M}\mathcal{A}$ . Here, the multigrid operator  $\mathcal{M}$  can be interpreted as a certain approximate inverse  $\mathcal{M} \approx \mathcal{A}^{-1}$ . It is not necessary to calculate this matrix explicitly; it is sufficient to evaluate the matrix-vector product  $\mathcal{M}\xi$ , i.e., to apply the multigrid iteration for a fixed right-hand side.

### 5.3.2 MULTIGRID AS A SOLVER

The multigrid iteration makes use of the hierarchy of finite element spaces

$$V_0 \subset V_1 \subset \dots \subset V_L,$$

obtained, for example, in the course of a systematic mesh refinement process; strategies for an automatic adaptive mesh refinement will be discussed below in Section 7. The connection between these spaces is given by “prolongation operators”  $P_{l-1}^l : V_{l-1} \rightarrow V_l$  and “restriction operators”  $R_l^{l-1} : V_l \rightarrow V_{l-1}$ . In the finite element context, these operators are given naturally as

$$P_{l-1}^l \text{ injection, } R_l^{l-1} \text{ } L^2 \text{ Projection.}$$

The main ingredients of a multigrid scheme are the smoothing operators  $S_l$  on each grid level  $0 \leq l \leq L$  ( $l=0$  corresponding to the coarse initial mesh and  $l=L$  to the finest mesh). The explicit form of these operators will be described below. The multigrid iteration

$$\mathcal{M}\xi = \mathcal{M}(l, z_0, \xi), \tag{7}$$

on level  $l$  with initial guess  $z_0$  and with  $m_1$  pre- and  $m_2$  post-smoothing steps is recursively defined as follows:

**Multigrid Algorithm**  $\mathcal{M}(l, z_0, \xi)$  for  $l \geq 0$ :

For  $l=0$ , the multigrid algorithm is given by an exact solver  $\mathcal{M}(l, z_0, \xi) := \mathcal{A}_0^{-1}\xi$ . For  $l>0$ , the following recursive iteration is performed:

1. Pre-smoothing  $m_1$  times:  $z_1 := S_l^{m_1} z_0$ .
2. Residual on level  $l$ :  $r_l := \xi - \mathcal{A}_k z_1$ .
3. Restriction to level  $l-1$ :  $r_{l-1} := R_l r_l$ .
4. Coarse grid correction starting with  $q_0 = 0$ :  $q := \mathcal{M}(l-1, q_0, r_{l-1})$ .
5. Prolongation to level  $l$ :  $z_2 := z_1 + P_{l-1} q$ .
6. Post-smoothing  $m_2$  times:  $\mathcal{M}(l, z_0, \xi) := S_l^{m_2} z_2$ .

If the multigrid recursion is applied  $\gamma$ -times on each mesh-level, one speaks of a  $V$ -cycle for  $\gamma = 1$  and of a  $W$ -cycle for  $\gamma = 2$ . If the multigrid iteration is used only as a preconditioner for a robust outer iteration scheme, usually the  $V$ -cycle suffices. If multigrid is used as the primary solver, particularly in the case of nonsymmetric problems, the  $W$ -cycle is more robust and therefore to be preferred. In this case, the  $F$ -cycle as indicated in the figure below is a compromise between  $V$ - and  $W$ -cycle. The multigrid cycle with  $\gamma > 2$  becomes too expensive and is not used.

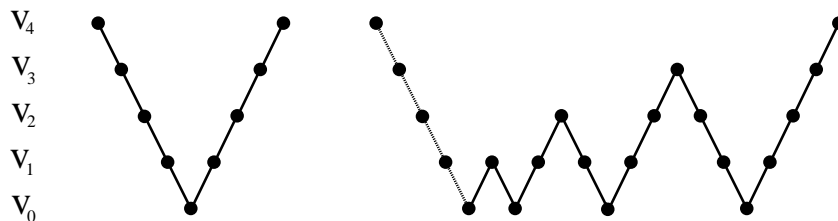


Figure 25: Scheme of the multigrid  $V$ -cycle (left) and the  $F$ -cycle (right)

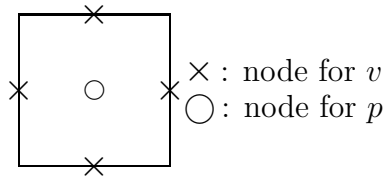
The design of a multigrid algorithm for solving the system (5) requires special care. In particular, the choice of the smoother is a delicate matter since the standard fixed-point iterations do not work for the indefinite matrix  $\mathcal{A}$ . This problem can be tackled in various ways.

(1) *Damped Jacobi smoother:* In the case  $\epsilon > 0$ , the matrix  $\mathcal{A}$  is weakly definite which makes it possible to apply even standard methods like the damped Jacobi iteration. However, the resulting algorithm is not very robust and parameter tuning requires care; see [38] for an application to 3-dimensional model problems. For larger Reynolds number the method slows down and multigrid convergence may get lost.

(2) *Block-Gauss-Seidel smoother:* A simple and successful smoother for the matrix  $\mathcal{A}$  can be obtained by a cell-wise blocking of the physical variables within a global Gauss-Seidel iteration. This was originally proposed by Vanka [101] for a finite difference discretization of the Navier-Stokes problem. We briefly discuss its analogue for the nonconforming “rotated”  $\tilde{Q}_1/P_0$  Stokes element. The velocity and pressure unknowns corresponding to a cell  $K$  or a patch of cells are grouped together. Indicating the corresponding element system matrices by index “ $loc$ ”, these blocks of local velocity and pressure unknowns are simultaneously updated according to

$$S_{loc}v_{loc}^{t+1} + B_{loc}p_{loc}^{t+1} = \text{“known”}, \quad B_{loc}^T v_{loc}^{t+1} = \text{“known”},$$

where  $S_{loc} = \sigma M_{loc} + \nu A_{loc} + N_{loc}(\bar{v}_h)$ . This iteration sweeps over all cell-blocks. The local Stokes problems have the dimension  $d_{loc} = 9$  (in 2D) or  $d_{loc} = 19$  (in 3D), respectively. The corresponding matrices (in 2D) are described in the following figure.



$$\mathcal{A}_{loc} = \begin{bmatrix} S_{loc,1} & O & B_{loc,1} \\ O & S_{loc,2} & B_{loc,2} \\ -B_{loc,1}^T & -B_{loc,2}^T & 0 \end{bmatrix}.$$

For cost reduction, the main diagonal blocks  $S_{loc,i}$  may be “lumped”,  $S_{loc,i} \approx D_{loc,i}$ . Furthermore, for increasing robustness, the iteration is damped,  $v_h^{t+1} = v_h^t + \omega(\tilde{v}_h^{t+1} - v_h^t)$  with  $\omega \approx 0.9$ . The good performance of this smoother for the  $\tilde{Q}_1/P_0$ -element has been demonstrated in Schreiber/Turek [83], Schieweck [82], and Turek [97], and for the  $Q_1/Q_1$ -element in Becker [7]. We illustrate the performance of the multigrid algorithm described above for the  $\tilde{Q}_1/P_0$ -element by results obtained for solving the driven-cavity problem on grids as shown below.

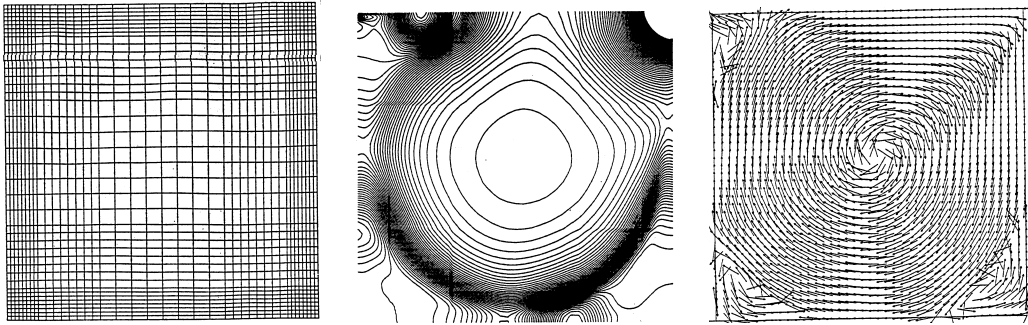


Figure 26: Driven cavity mesh (left) and computed results: pressure isolines (middle), velocity plot (right).

Table 2: Multigrid convergence rates (2 pre- and 1 post-smoothing step by the “Vanka smoother”) and number of outer fixed-point iterations on uniformly refined meshes.

#cells	1600	6400	25600	#iter
Re = 1	0.081	0.096	0.121	4
Re = 100	0.098	0.099	0.130	6
Re = 1000	0.227	0.245	0.168	9
Re = 5000	0.285	0.368	0.370	18

We note that a similar block-iteration can also be used in the context of a *incomplete* block-LU-decomposition for generating a multigrid smoother; for a detailed discussion of this approach see Braack [17].

From the common multigrid theory for elliptic equations we know that point iterations loose the smoothing property for large mesh-aspect ratios  $\sigma_h$ . The remedy is the use of a smoother which becomes a direct solver in the limit  $\sigma_h \rightarrow \infty$ . Consequently, since our smoother acts like a point–Gauss–Seidel

iteration on the velocity unknowns, we expect problems in the case of strongly stretched grids. Our strategy to overcome this difficulty is as follows: Since we expect the cell aspect ratio  $\sigma_K$  to be large only in a small part of the computational domain, we should use an *adaptive* smoother. This means that we will combine the point smoother with a more robust version just where we need it, for instance on elements with large  $\sigma_K$ . In this approach the nodes are grouped in the direction of the anisotropic mesh refinement and iterated implicitly leading to a process which may be termed “stringwise” block-Gauss-Seidel method.

Let us finally mention a critical problem especially in the use of non-uniform grids. The use of iterative solvers makes it necessary to define a stopping criterion. To this end, we need to measure the residual in the right norm. Clearly, the common weighting by the number of unknowns is not appropriate on non-uniform grids. For an approach towards a solution of this problem based on the Galerkin orthogonality inherent to the multigrid process, we refer to [11] and Becker [8].

(3) *Discrete projection smoother*: Finally, we present an approach to constructing multigrid solvers for the indefinite system (5) which uses the idea of operator splitting as introduced above in Section 4 on time-discretization schemes; see Turek [95, 97]. This method is particularly efficient in the nonstationary case when  $\sigma = 1/k$  balances  $\nu/h^2$ . In the following, we consider the linearized problem arising within a time-stepping scheme as described above in combination with spatial discretization by a Stokes element which does not need pressure stabilization. This problem has the form

$$Sv^n + Bp^n = g^n, \quad B^T v^n = 0, \quad (8)$$

with the (momentum) matrix  $S = \sigma M + \nu A + N(\bar{v}^n)$ . The right-hand side  $g^n$  and the approximation  $\bar{v}^n$  are given from the preceding time level. Elimination of the velocity unknown yields again the Schur complement formulation

$$B^T S^{-1} B p^n = B^T S^{-1} g^n, \quad v^n = S^{-1}(g^n - B p^n). \quad (9)$$

We have already mentioned that the solution of this problem by Krylov space methods with evaluation of  $S^{-1}$  by multigrid iteration becomes increasingly inefficient for small time step  $k$ , larger Reynolds number, and on strongly anisotropic meshes. This problem can be overcome by using instead a simple Richardson iteration for the Schur complement equation (9) with a preconditioner of the form  $B^T C^{-1} B$ . Popular choices for the preconditioning operator  $C$  are:

- $C^{-1} = I$  (corresponds to the SIMPLE algorithm).
- $C^{-1} = \bar{M}^{-1}$  (lumped mass preconditioning).
- $C^{-1} = \bar{M}^{-1} + \alpha^{-1} B^T B$  (Turek’s preconditioner)

The resulting iteration is termed “discrete projection method” (see Turek [95]):

$$p^{n,l+1} = p^{n,l} - (B^T C^{-1} B)^{-1} (B^T S^{-1} B p^{n,l} - B^T S^{-1} g^n). \quad (10)$$

After  $L$  iteration steps, on sets  $p^n := p^{n,L}$  and computes the corresponding velocity component by solving:

$$Sv^n = g^n - Bp^n + \alpha^{-1}(\alpha I - Sc^{-1})B(p^{n,L} - p^{n,L-1}),$$

with some relaxation parameter  $\alpha \in (0, 1)$ . This construction of  $v^n$  ensures that the resulting velocity is in the discrete sense divergence-free,  $B^T v^n = 0$ , and suggests the name “projection method” for the whole scheme. The discrete projection method is then used as a smoother within an outer multigrid iteration.

In the special case  $L = 1$ , this scheme corresponds to a discrete version of the classical projection methods of Chorin (for the choice  $p^{n,0} := 0$ ) and of Van Kan (for the choice  $p^{n,0} := p^{n-1}$ , see Gresho [32]). This operator-splitting time-stepping scheme has the form:

1.  $S\tilde{v}^n = g^n - kBp^{n-1}$  (Burgers step),
2.  $B^T \bar{M}^{-1} Bq^n = k^{-1} B^T \tilde{v}^n$  (Pressure Poisson equation),
3.  $v^n = \tilde{v}^n - k\bar{M}^{-1} Bq^n$  (Velocity update),
4.  $p^n = p^{n-1} + \alpha q^n$  (Pressure update).

All these schemes are variants of the “segregated” solution approach containing the schemes of SIMPLE-type and other pressure correction schemes as special cases; for a survey see [97] and [98].

The multigrid method with smoothing by the discrete projection iteration (10) has proven to be a very efficient solution method for the fully coupled problem (8); it is robust for all relevant Reynolds number (laminar flows) and time steps. The whole solution process is based on efficient and robust “inner” multigrid solvers for the subproblems “Burgers equation” and “pressure Poisson equation”. The concrete implementation of this algorithm (as described in Turek [97]) requires about 1 KByte memory per mesh cell and shows almost meshsize-independent convergence behavior. As the result, 3D simulations with more than  $10^7$  unknowns requiring about 1 GByte of memory can be done on modern workstations.

**Open Problem 5.2:** *Derive a good preconditioner (smoother) for the Schur complement iteration (10) in the transport-dominant case.*

## 6 A review of theoretical analysis

In this section, we give an account of the available theoretical analysis for the discretization described in the previous sections. We concentrate on the practical impact of these theoretical results; the main topics are:

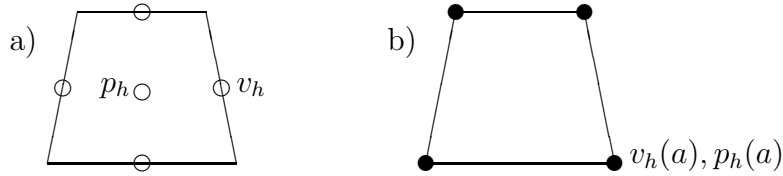
- Problem of regularity at “ $t = 0$ ”.
- Problem of global convergence up to “ $t = \infty$ ”.
- Problem of realistic error constants.

We will identify some critical shortcomings of the available theory which lead to challenging questions for further analysis.

We assume that the stationary or nonstationary Navier-Stokes equations are discretized by the finite element method as specified in Section 3 combined with one of the time-stepping schemes described in Section 4.

(I) For the spatial discretization, we recall the following two representative examples of (quadrilateral) Stokes elements:

- a) the nonconforming “rotated”  $d$ -linear  $\tilde{Q}_1/P_0$  element;
- b) the conforming  $d$ -linear  $Q_1/Q_1$  element with pressure stabilization.



These discretizations are of second order expressed in terms of local approximation properties of the finite element functions used:

$$\inf_{\phi_h \in \mathbf{H}_h} \|v - \phi_h\| \leq c h^2 \|\nabla^2 \phi\|, \quad \phi \in \mathbf{H} \cap \mathbf{H}^2(\Omega).$$

(II) For the time discretization, we think of the Crank-Nicolson scheme or the Fractional-Step- $\theta$  scheme which are both of second order in terms of local truncation error, e.g., for the Crank-Nicolson scheme applied to the homogeneous heat equation, there holds

$$\|k^{-1}(v^n - v^{n-1}) + \frac{1}{2}(A^n v^n + A^{n-1} v^{n-1})\| \leq c k^2 \max_{[t_{n-1}, t_n]} \|\partial_t^2 v\|_{-1}.$$

In view of these *local* approximation properties and the stability of the schemes, we expect a *global* a priori estimate for the errors  $e_v^n := u(\cdot, t_n) - u_h^n$  and  $e_p^n := p(\cdot, t_n) - p_h^n$  of the form

$$\max_{0 < t_n \leq T} \{\|e_v^n\| + \|e_p^n\|_{-1}\} \leq C(\nu, T, \text{data})\{h^2 + k^2\}, \quad (1)$$

with an “error constant”  $C(\nu, T, \text{data})$  depending on the viscosity parameter  $\nu$ , the time-interval length  $T > 0$ , and assumed bounds  $M$  for the data of the problem, e.g.,

$$M := \|\nabla^2 v^0\| + \sup_{[0, T]} \{\|f\| + \|\partial_t f\|\} < \infty.$$

If additionally the domain  $\Omega$  is sufficiently regular (say, convex or with  $C^2$ -boundary), it is guaranteed that the solution  $\{v, p\}$  satisfies at least the a priori estimate

$$\sup_{(0, T]} \{\|\nabla^2 v\| + \|\partial_t v\| + \|\nabla p\|\} < \infty. \quad (2)$$

Clearly, the size of the error constant  $C(\nu, T, \text{data})$  is of crucial importance for the practical value of the error estimate; we will come to this point in more detail, below. At first, we have to consider the question of whether an error estimate of the form (1) can be expected to hold at all. In general, the answer is “no”, unless certain additional conditions are satisfied. This leads us to the following discussion of the “smoothing property”.

### 6.1 The problem of regularity at “ $t = 0$ ”

The second-order convergence of the time-stepping scheme expressed in the estimate (1) requires an a priori bound of the form  $\sup_{(0, T]} \|\partial_t^2 v\| < \infty$ . We have seen in Section 2 that there is a principle problem with assuming this degree of regularity in general. Even for arbitrarily smooth data the solution of the Navier-Stokes problem may suffer from

$$\lim_{t \rightarrow 0} \{\|\nabla^3 v(t)\| + \|\nabla \partial_t v(t)\|\} = \infty, \quad (3)$$

unless certain *non-local* (and non-verifiable) compatibility conditions are satisfied for the initial data. We recall from Section 2 the natural regularity assumption for the (nonstationary) Navier-Stokes equations (without additional compatibility condition):

$$v^0 \in \mathbf{J}_1(\Omega) \cap \mathbf{H}^2(\Omega) \quad \Rightarrow \quad \sup_{t \in (0, T]} \{\|\nabla^2 v(t)\| + \|\partial_t v(t)\|\} < \infty. \quad (4)$$

Accordingly, the best possible error estimate for the velocity which can be obtained under these “realistic” assumptions is

$$\sup_{t_n \in (0, T]} \|e_v^n\| = \mathcal{O}\{h^2 + k\}. \quad (5)$$

This estimate is only of first order in time, in contrast to the postulated second-order error estimate (1). As a result of the foregoing discussion we obtain the following:

**Conclusion:** *For any discretization of the nonstationary Navier-Stokes equations which requires more than the natural regularity inherent to the problem, meaningful higher-order error estimates must be of “smoothing type”.*

We call an error estimate of type (1) a “smoothing error estimate” if it is of the form

$$\sup_{t_n \in (0, T]} \{t_n \|e_v^n\| + t_n^{3/2} \|e_p^n\|_{-1}\} \leq C(\nu, T, \text{data})\{h^2 + k^2\}. \quad (6)$$

This estimate reflects the well-known “smoothing behavior” of the exact solution  $\{v, p\}$  as  $t \rightarrow 0$  in the (realistic) situation (4):

$$\sup_{t \in (0, T]} \{t^{r/2-1} \|\nabla^r v(t)\| + t^{r-1} \|\partial_t^r v(t)\|\} \leq c\{\|\nabla^2 v^0\| + \text{data}\}. \quad (7)$$

Smoothing error estimates of the form (6) have been established earlier for standard parabolic problems like the heat equation in the case of rough initial data; see, e.g., Thomée [89], as well as [62] and [71]. Corresponding results for the Navier-Stokes equation have been given in [43] for higher-order spatial semi-discretization and in [44] for the Crank-Nicolson time-stepping scheme. It turns out that due to the nonlinearity of the problem, the maximal achievable orders of smoothing error estimates under assumption (4) is  $\mathcal{O}(h^6)$  for the spatial discretization and accordingly  $\mathcal{O}(k^3)$  for the time stepping (provided that the scheme is strongly A-stable). This particularly implies the result (6) stated above. The existence of a natural order-barrier for the smoothing property of finite element Galerkin schemes applied to nonlinear problems has been established by Johnson, et al. [53]. We adapt the following example from [53] for the situation of  $H^2$ -regular initial data as relevant for the case of the nonlinear Navier-Stokes equations.

**Example:** *Example of limited smoothing property*

For  $x \in (-\pi, \pi)$  and  $t > 0$ , we consider the system of equations

$$\begin{aligned} \partial_t u - \partial_x^2 u &= 4 \min\{v^2, 1\}, & u(x, 0) &= u^0(x) := 0, \\ \partial_t v - \partial_x^2 v &= 0, & v(x, 0) &= v^0(x) := m^{-r} \cos(mx), \end{aligned}$$

with periodic boundary conditions. For any fixed  $m \in \mathbb{N}$  and  $r \in \mathbb{N} \cup \{0\}$ , the exact solution is

$$\begin{aligned} u(x, t) &= m^{-2r-2} (1 - e^{2m^2 t}) (1 + e^{2m^2 t} \cos(2mx)), \\ v(x, t) &= m^{-r} e^{-m^2 t} \cos(mx). \end{aligned}$$

For spatial semi-discretization of this problem, let the Galerkin method be used with the trial spaces

$$S_m := \text{span}\{1, \cos(x), \sin(x), \dots, \cos((m-1)x), \sin((m-1)x)\},$$



and let  $P_m$  denote the  $L^2$  projection onto  $S_m$ . Since  $P_m v^0 = 0$ , taking as usual  $P_m u^0$  and  $P_m v^0$  as initial values for the Galerkin approximation results in the Galerkin solutions  $v_m(t) = 0$  and  $u_m(t) = 0$ . Consequently, for fixed  $t > 0$ , there holds

$$\|(u_m - u)(t)\| = \|u(t)\| \sim \sqrt{2\pi} m^{-2r-2} = \sqrt{2} \|v^0\|_r h^{2r+2},$$

if we set  $h := m^{-1}$ . This demonstrates that, for  $v^0 \in H^2(-\pi, \pi)$ , i.e., for  $r = 2$ , the best achievable order of approximation for  $t > 0$  is indeed  $\mathcal{O}(h^6)$ .

There is another remarkable aspect of the estimate (6) which concerns the Crank-Nicolson scheme. This scheme, due to its absent damping properties (not *strongly* A-stable), possesses only a reduced smoothing property. In consequence, even in the case of the linear heat equation, for initial data  $v^0 \in L^2(\Omega)$  only qualitative convergence  $\|e_v^n\| \rightarrow 0$  ( $h, k \rightarrow 0$ ) can be guaranteed at fixed  $t_n = t > 0$ . For even stronger initial irregularity (e.g.,  $v^0 = \delta_x$  a Dirac measure) divergence  $\|e^n\| \rightarrow \infty$  ( $h, k \rightarrow 0$ ) occurs. However, the optimal smoothing behavior is recovered if one keeps the relation  $k \sim h^2$ . This undesirable step-size restriction can be avoided simply by starting the computation with a few (two or three) backward Euler steps; for examples and an analysis, see [61], [71], and the literature cited therein. Surprisingly, such a modification is not necessary for more regular initial data (half way up to the maximum regularity),  $v^0 \in H_0^1(\Omega) \cap H^2(\Omega)$ . In this case the Crank-Nicolson scheme admits an optimal-order smoothing error estimate of the form

$$\|e_v^n\| \leq C(\nu, T, \|\nabla^2 v^0\|, \text{data}) \{h^2 + t_n^{-1} k^2\}, \quad t_n \in (0, T]. \quad (8)$$

For the heat equation this is easily seen by a standard spectral argument; see Chen/Thomé [23] and [71]. The extension of the smoothing error estimate (8) to the nonlinear (nonsymmetric and nonautonomous) Navier-Stokes equations is one of the main results in [44].

Finally, we mention some further results on smoothing error estimates relevant for the Navier-Stokes equations together with some open problems:

- Second-order projection schemes, particularly the Van Kan scheme, have been analyzed by Prohl [69, 70] and optimal-order smoothing error estimates have been established.
- The second-order smoothing property of the Fractional-Step- $\theta$  scheme has been proven by Müller [64].

**Open Problem 6.1:** *The construction of compatible initial data from experimental data has already been formulated as a problem. If this is not possible, it would be interesting to estimate the time length over which the error pollution effect of incompatible initial data persists.*

**Open Problem 6.2:** *Establish the optimal smoothing property of any higher-order ( $q \geq 3$ ) time discretization schemes for the Navier-Stokes equations.*

## 6.2 The problem of convergence up to “ $t = \infty$ ”

The error constants in the a priori error estimates (1) usually grow exponentially in time,

$$C(\nu, T, \text{data}) \sim Ke^{\kappa T},$$

unless the data of the problem is very small, actually of size  $\nu^2$ , such that nonlinear perturbation terms can be absorbed into the linear main part. This exponential growth is unavoidable in general, due to the use of Gronwall’s inequality in the proof. In fact, the solution to be computed may be exponentially unstable, so that a better error behavior cannot be expected. To improve on this situation, one has to make additional assumptions on the stability of the solution. Instead of requiring the data of the problem to be unrealistically small, the solution  $\{v, p\}$  itself is supposed to be stable. This assumption may not be verifiable theoretically; nevertheless it may be justified in many situations in view of experimental evidence. A discussion of various types of stability concepts for nonstationary solutions of the Navier-Stokes equations in view of numerical approximation can be found in [42, 45].

*Exponential Stability:* A solenoidal solution  $v$  is called (conditionally) “exponentially stable”, if for each sufficiently small initial perturbation  $w^0 \in \mathbf{J}_1(\Omega)$ ,  $\|w^0\| < \delta$ , at any time  $t_0 \geq 0$ , the solution  $\tilde{v}(t)$  of the perturbed problem

$$\partial_t \tilde{v} - \nu \Delta \tilde{v} + \tilde{v} \cdot \nabla \tilde{v} + \nabla q = 0, \quad t \geq t_0,$$

starting from  $\tilde{v}(t_0) = v(t_0) + w^0$ , satisfies

$$\|(v - \tilde{v})(t)\| \leq Ae^{-\alpha(t-t_0)}\|w^0\|, \quad t \geq t_0, \quad (9)$$

with certain constants  $A > 0$  and  $\alpha > 0$ . In this assumption it is essential that the decay of the perturbation is proportional to the size of the initial perturbation  $\|w^0\|$ . For global strong solutions this concept of exponential “ $L^2$ -stability” is equivalent to corresponding stability concepts expressed in terms of stronger norms, e.g. the  $H^1$  norm; see [45]. It has been proved in a series of papers [42, 43, 44] that exponentially stable solutions can be approximated uniformly in time, i.e.,

$$\sup_{t_n \geq 1} \|v_h^n - v(\cdot, t_n)\| \leq C\{h^2 + k^2\}. \quad (10)$$

In this estimate the error constant  $C = C(A, \alpha)$  depends on the stability parameters of the solution. The proof uses a continuation argument. We sketch its essential steps for semi-discretization in time by the backward Euler scheme.

*Proof of the global error estimate (10):*

(i) We recall the *local* bound for the error  $e^n = v(t_n) - v_k^n$ ,

$$\|e^n\| \leq E(t_n)k, \quad t_n \geq 0, \quad (11)$$

involving the exponentially growing error constant  $E(t) := Ke^{\alpha t}$ . By continuity, it can be assumed that this error estimate holds with the same constant for all solutions neighboring the true solution  $v$ . The proof of this statement is technical and uses the particular properties of the discretization scheme considered. Further, let  $T = Nk$  be a fixed time length such that with the stability parameters of the solution  $v$ , there holds

$$Ae^{-\alpha T} \leq \frac{1}{2}. \quad (12)$$

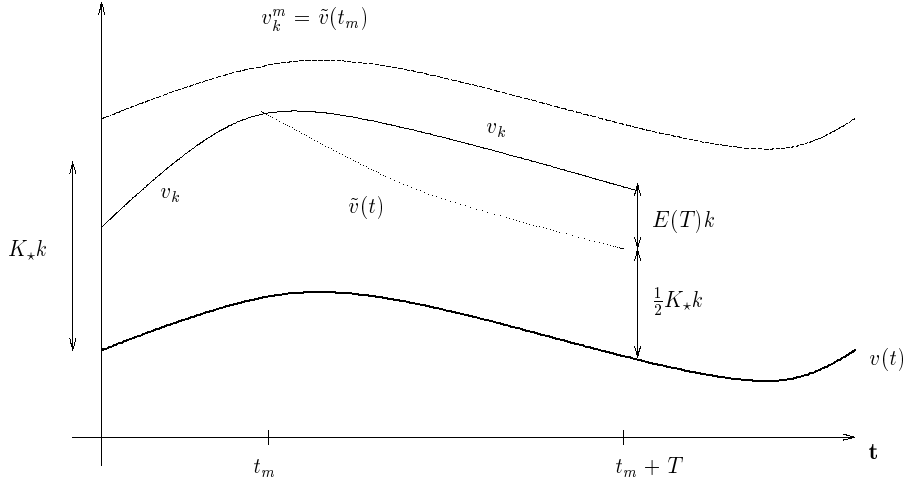


Figure 27: Scheme of induction proof (following [42])

(ii) Suppose now that the desired error estimate is already known to hold on some time interval  $(0, t_m]$ ,  $t_m \geq T$ , with error constant  $K_* := 2E(T)$ . Let  $\tilde{v}(t)$  be the solution of the perturbed problem

$$\partial_t \tilde{v} - \nu \Delta \tilde{v} + \tilde{v} \cdot \nabla \tilde{v} + \nabla \tilde{p} = 0, \quad t \geq t_m,$$

starting at  $t_m$  with initial value  $\tilde{v}(t_m) = v_k^m$ . In virtue of the assumed exponential stability of the solution, there holds

$$\|(v - \tilde{v})(t)\| \leq Ae^{-\alpha(t-t_m)} \|e^m\|, \quad t > t_m,$$

for sufficiently small  $k$  guaranteeing  $\|e^m\| < \delta$ . Then, stepping forward by time length  $T$ , we obtain

$$\|e^{m+n}\| \leq \|(v - \tilde{v})(t_m + T)\| + \|\tilde{v}(t_m + T) - v_k^{m+n}\|.$$

Here, the first term is bounded by  $K_*k$ , in view of (11) and the induction assumption (12), while the second one can be controlled by  $E(T)k$ , using the local error estimate (11) for  $\tilde{v} - v_k$ . Hence, it follows that

$$\|e^{m+n}\| \leq K_*k.$$

The assertion then follows by induction with respect to multiples of  $T$ .

The argument presented for the global error estimate (10) appears simple and general; however, in concrete situations involving simultaneous discretization in space and time there are several technical difficulties. The initial value for the perturbed solution  $\tilde{v}(t_n)$  in the induction step may not be admissible (i.e., not exactly divergence-free or even nonconforming). Further, the use of the local error bound (11) for the perturbed error  $\tilde{v} - v_h$  requires control on higher-order regularity of the corresponding initial value  $\tilde{v}(t_n)$ . These and some other complications can be overcome as shown in [42, 43, 44], for different types of spatial as well as time discretization.

### 6.3 The problem of realistic error constants

In the preceding sections, we have discussed the derivation of *qualitative* a priori error estimates, local as well as global in time. Now, we turn to the more *quantitative* aspect of the size of error constants relating to the question of practical relevance of the a priori results. To this end, let us briefly summarize the results of a priori error analysis presented so far:

a) In the *stationary* case, we can guarantee convergence behavior like

$$\|e_v\| \leq Ch^p$$

provided that the solution  $v$  is sufficiently smooth and locally unique (i.e., the linearization of the nonlinear Navier-Stokes operator at  $v$  is regular). Then, the error constant  $C(\nu, v)$  depends on bounds on the regularity of  $v$  as well as its “stability”, on the viscosity  $\nu$ , and of course on the characteristics of the discretization.

b) In the *nonstationary* case, we can guarantee convergence behavior like

$$\|e_v^n\| \leq C\{h^p + k^q\} \quad 0 \leq t_n \leq T,$$

provided again that the solution  $v$  is sufficiently smooth. The error constant  $C(\nu, T, v)$  depends on bounds on the regularity of  $v$ , on the viscosity  $\nu$ , and additionally on the length of the time interval  $T$ .

A question naturally arises: How large is  $C$ ? In “normal” situations as, for example, for the Poisson problem or the heat equation, the error constant may be shown to be of moderate size  $C \sim 1 - 10^4$ , depending on the situation and the care spent in the estimation. The qualitative conclusion from the estimates

may then be that the error bound is reduced by a factor of  $2^{-\min\{p,q\}}$  if the mesh size is halved in space and time. The (not unrealistic) hope is that this carries over to the true discretization error. Unfortunately, the Navier-Stokes equations do not at all a “normal” problem; it is of mixed elliptic-hyperbolic or parabolic-hyperbolic type with degenerating ellipticity. This has decisive consequences for the size of the error constants  $C$ .

Normalizing the flow configuration as usual to characteristic length  $L = 1$  and velocity  $U = 1$ , the Reynolds number for common cases is  $Re = \nu^{-1} \approx 1 - 10^5$ , which relates to “laminar” flow, and the characteristic time length is  $T \sim 1/\nu$ . This means that it takes the time  $T \approx \nu^{-1}$  for the flow to reach a characteristic limit behavior, e.g. stationary or time periodic. The question can now be made more precise: *How do the error constants  $C$  depend on  $Re$ ?* This dependence has several sources:

- the explicit occurrence of  $\nu$  in the differential operator,
- the dependence of the solution’s regularity on  $\nu$  (boundary layers),
- the dependence on the length of the time interval  $T \sim 1/\nu$ ,
- the dependence of the solution’s stability on  $\nu$ .

Let us discuss the mechanisms of these dependencies separately.

*(i) Structure of the differential operator:*

The standard procedure in the stationary case is to absorb the lower-order terms into the linear main part  $-\nu\Delta v$  which leads to the dependence

$$C \sim \nu^{-1}.$$

In the nonstationary case the lower-order terms are absorbed into the term  $\partial_t v$  by the use of Gronwall’s inequality resulting in

$$C \sim e^{KT/\nu}, \quad K \sim \sup_{(0,T]} \|\nabla v\| \sim \nu^{-1/2}.$$

This dependence on  $\nu$  can be formally removed by using streamline-diffusion damping for the transport term, but it leaves the  $T$ -dependence.

*(ii) Regularity of solution:*

For small  $\nu$  boundary layers of width  $\delta \sim \sqrt{\nu}$  occur. This implies that

$$\sup_{\Omega} |\nabla v| \sim \nu^{-1/2}, \quad C \sim \|\nabla^p v\| \sim \nu^{-\alpha(d,p)}.$$

This problem can be solved by proper mesh refinement in the boundary layer.

*(iii) Length of the time interval:*

It was demonstrated above that the local “worst case” error constant

$$C \sim e^{KT/\nu}$$

becomes independent of the time interval-length  $T$ ,

$$C \sim e^{KT_*/\nu},$$

if the solution can be assumed to be exponentially stable. Here,  $T_*$  is sufficiently large but fixed. However, tracing constants in the proof, we see that  $T_* \sim \nu^{-1}$  rendering this formally global error bound practically meaningless for small  $\nu$ .

*(iv) Stability of the solution:*

The argument for proving error estimates for nonlinear problems rely on assumptions on the stability of certain linearized tangent operators. The resulting error constants can in general not be assumed to behave better than

$$C \sim e^{KT/\nu}.$$

This exponential dependence seems unavoidable unless something different is shown in particular situations. The observability of laminar flows even for higher Reynolds numbers indicates that these flows may possess better stability properties than expressed by the “worst case” scenario addressed above. An analogous conclusion may be true even for turbulent flows with respect to certain averaged quantities.

**Conclusion:** *In general, one has to admit that the error constants depend exponentially on  $\nu^{-1}$ , unless something different is proven. Realizing that even in the range of laminar flows,  $20 \leq Re \leq 10^4$ ,*

$$e^{20} \approx 5 \cdot 10^8, \quad e^{100} \approx 10^{43}, \quad e^{1000} \approx \infty,$$

*the practical meaning of available a priori error estimates seems rather questionable!*

The above observation seems to indicate that there is a conceptual crisis in the theoretical support of CFD as far as it concerns the computational solution of the Navier-Stokes equations. This is contrasted by the abundant body of research papers reporting successful computations of viscous flows and the good agreement of the obtained results with experimental data. Hence, we reformulate the question: *Is there any theoretical support that certain flows (i.e., solutions of the Navier-Stokes equations) can actually be computed numerically.* If the answer were “no”, everybody should be worried. We again emphasize that the presence of an asymptotic error estimate of the form

$$\|v - v_h\| \leq C\{h^p + k^q\}$$

cannot be taken as justification for the meaningful performance of a numerical scheme, unless the error constant  $C$  is shown to be of moderate size at least for certain model situations of practical interest. Reliable flow simulation requires computable error bounds in terms of the approximate solution; the elements of such an *a posteriori* error analysis will be described in Section 7 below.

In proving useful error estimates, we have to deal with the question of proper concepts for describing the stability of solutions relevant for numerical approximation. Qualitatively, all stability concepts may be equivalent but this strongly depends on the viscosity parameter  $\nu$ . The choice of the wrong norm may lead to unfavorable dependence on  $\nu$ , like  $\mathcal{O}(\nu^{-2})$  rather than the generic behavior  $\mathcal{O}(\nu^{-1})$ . Actually, the fundamental question whether there are practically interesting situations in which the solution of the Navier-Stokes equations are stable, with stability constant  $c_S \sim \nu^{-1}$  seems open. Results in this direction appear necessary for a rigorous error analysis of discretization schemes. However, until now, practically meaningful a priori error bounds are not even available for such basic situations as Couette flow (constant shear flow) and Poiseuille flow (constant pipe flow); we will address this question in more detail in the following section.

#### 6.4 Towards a “quantitative” a priori error analysis

The following discussion is of conceptual nature. In order to abstract from the nonessential technicalities of finite element discretization, we consider the idealized situation of an “exactly” divergence-free approximation, using subspaces  $\mathbf{V}_h \subset \mathbf{V} := \mathbf{J}_1(\Omega)$ . Accordingly, the discretization delivers only approximations  $v_h \in \mathbf{V}_h$  to the velocity  $v \in \mathbf{V}$ . The associated pressures  $p_h$  are then to be determined by post-processing. Further, we restrict us to the very basic case of homogeneous Dirichlet boundary conditions  $v|_{\partial\Omega} = 0$ .

##### 6.4.1 THE STATIONARY CASE

We begin with the stationary Navier-Stokes problem

$$-\nu\Delta v + v \cdot \nabla v + \nabla p = f, \quad \nabla \cdot v = 0, \quad \text{in } \Omega, \quad v|_{\partial\Omega} = 0. \quad (13)$$

Using again the notation

$$a(v, \psi) = \nu(\nabla v, \nabla \psi), \quad n(v, v, \psi) = (v \cdot \nabla v, \psi),$$

the “pressure-free” variational formulation seeks  $v \in \mathbf{V}$ , such that

$$A(v; \phi) := a(v, \phi) + n(v, v, \phi) = (f, \phi) \quad \forall \phi \in \mathbf{V}. \quad (14)$$

The corresponding finite element discretization seeks  $v_h \in \mathbf{V}_h$ , such that

$$A(v_h; \phi_h) = (f, \phi_h) \quad \forall \phi_h \in \mathbf{V}_h. \quad (15)$$

All error analysis of this discretization is based upon the (nonlinear) Galerkin orthogonality:

$$A(v; \phi_h) - A(v_h; \phi_h) = 0, \quad \phi_h \in \mathbf{V}_h. \quad (16)$$

In the following, we denote the error by  $e := v - v_h$ .

a) The “small data case”,  $\|\nabla v\| \sim \nu$ :

The Fréchet derivative taken at  $v$  of the semi-linear Form  $A(\cdot; \cdot)$  is given by

$$L(v; \phi, \psi) = a(\phi, \psi) + n(v, \phi, \psi) + n(\phi, v, \psi).$$

Under the “small data” assumption, this bilinear form is coercive on  $\mathbf{V}$  with “stability constant”  $c_S = c_S(\nu) \sim \nu^{-1}$ :

$$\|\nabla \phi\| \leq c_S L(v; \phi, \phi).$$

Linearization and Galerkin orthogonality then leads to the relation

$$\|\nabla e\| \leq c_S L(v, e, e) = c_S \{n(e, e, e) + A(v; v - \phi_h) - A(v_h; v - \phi_h)\},$$

with an arbitrary approximation  $\phi_h \in \mathbf{V}_h$  to  $v$ . From this we infer that, for sufficiently small  $h$ ,

$$\|\nabla e\| \leq c_S C h, \quad (17)$$

with an error constant  $C = C(\|\nabla^2 v\|, \text{data})$ .

b) The general case of an “isolated” solution:

Now, the solution  $v$  is assumed to be stable in the sense that

$$\|\nabla \phi\| \leq c_S \sup_{\psi \in \mathbf{V}} \frac{L(v, \phi, \psi)}{\|\nabla \psi\|}, \quad (18)$$

with some “stability constant”  $c_S = c_S(\nu, v)$ . Again, by linearization and Galerkin orthogonality, it follows that

$$\|\nabla e\| \leq c_S C h. \quad (19)$$

Further, assuming stability of the Fréchet derivative in the form

$$\|\phi\| \leq c_S \sup_{\psi \in \mathbf{V} \cap \mathbf{H}^2(\Omega)} \frac{L(v, \phi, \psi)}{\|\nabla^2 \psi\|}, \quad (20)$$

one may apply the usual duality argument to obtain the  $L^2$ -error bound

$$\|e\| \leq c_S C h^2. \quad (21)$$

These results rely on the assumption of stability of the problem expressed in terms of the stability constant  $c_S$ ; see [56]. In order to use the resulting estimates, one has to determine these constants either analytically, which may rarely be possible, or computationally by solving dual problems. Numerical experiments for the driven-cavity problem reported by Boman [16] show a dependence like  $c_S(\nu) \sim \nu^{-1}$  in the (“laminar”) range  $1 \leq Re \leq 10^3$ . This investigation should be extended to other elementary flows in order to see whether linear growth  $c_S \sim Re$  is generic for laminar flow. The answer is not clear yet, as indicated by the following simple example.



An example of “bad” stability (from Tobiska/Verfürth [91]): For the worst-case scenario, we quote the one-dimensional Burgers equation

$$-\nu v_{xx} + vv_x = 0, \quad x \in (-1, 1), \quad v(-1) = 1, \quad v(1) = -1.$$

The exact solution is  $v(x) = -2\alpha_\nu \nu \tanh(\alpha_\nu x)$ , where  $\alpha_\nu$  is the unique positive solution of  $2\nu\alpha_\nu \tanh(\alpha_\nu) = 1$ . Linearization at this solution results in the boundary value problem

$$-\nu z_{xx} + vz_x + zv_x = f \quad x \in (-1, 1), \quad z(-1) = z(1) = 0,$$

which has the solution

$$z(x) = -\nu^{-1} e^{U(x)/\nu} \int_{-1}^x e^{-U(t)/\nu} \left( \int_0^t f(s) ds + c \right) dt,$$

where  $U$  is a primitive of  $v$  and the constant  $c$  is determined by imposing the boundary condition  $z(1) = 0$ . We ask for the best possible bound in the stability estimate

$$\nu \|z\|_{H^1} \leq C(\nu) \|f\|_{H^{-1}}.$$

For the particular choice  $f(x) = \cosh(\alpha_\nu x)$ , there holds

$$z(x) = \frac{\cosh^3(\alpha_\nu) - \cosh^3(\alpha_\nu x)}{3\alpha_\nu^2 \nu \cosh^2(\alpha_\nu x)}.$$

Since

$$\|z\|_\infty \leq \sqrt{2} \|z\|_1, \quad \|f\|_{-1} \leq 2\sqrt{2} \|f\|_\infty,$$

it follows that, for  $\nu \leq 1$ ,

$$\|z\|_\infty \geq c\nu e^{1/\nu} \|f\|_\infty,$$

and consequently,  $C(\nu) \sim e^{1/\nu}$ . This seems to indicate that Burgers equation is not numerically solvable which, however, contradicts practical evidence. The explanation may be that for the performance of discretization stability is essential in other more local measures than those considered above.

**Open Problem 6.3:** *Explain the success of discretization methods in computing solutions to the Burgers equation despite its bad conditioning with respect to the “energy norm”.*

#### 6.4.2 THE NONSTATIONARY CASE

We now turn to the nonstationary Navier-Stokes problem posed on a time interval  $I = [0, T]$ ,

$$\partial_t v - \nu \Delta v + v \cdot \nabla v + \nabla p = f \quad \text{in } \Omega \times I, \quad v|_{\partial\Omega} = 0, \quad v|_{t=0} = v^0. \quad (22)$$

The corresponding (pressure-free) space-time variational formulation uses the function space  $\mathbf{V}(I) := H^1(I; \mathbf{V})$  and the space-time forms

$$\begin{aligned}(\phi, \psi)_I &= \int_I (\phi, \psi) dt, & a_I(\phi, \psi) &= \int_I a(\phi, \psi) dt, \\ n_I(v, \phi, \psi) &= \int_I n(v, \phi, \psi) dt.\end{aligned}$$

Then, a solution  $v \in \mathbf{V}(I)$  is sought satisfying

$$A(v; \phi) = F(\phi) \quad \forall \phi \in \mathbf{V}(I), \quad (23)$$

where  $F(\phi) := (f, \phi)_I + (v^0, \phi(0))$  and

$$A(v; \phi) := (\phi, \psi)_I + a_I(\phi, \psi) + n_I(v, \phi, \psi) + (v^0, \phi(0)).$$

Here, the initial condition  $v(0) = v^0$  is incorporated weakly.

In the following conceptual discussion, we restrict ourselves to the semi-discretization in time leaving the spatial variable “continuous”. The discretization is by the “discontinuous” Galerkin method of degree  $r = 0$  (“dG(0) method”) which is a variant of the backward Euler scheme. The time interval  $I = [0, T]$  is decomposed like  $0 = t_0 < t_1 < \dots < t_{M+1} = T$ , and we set

$$I_m = [t_{m-1}, t_m), \quad k(t) \equiv k|_{I_m} = t_m - t_{m-1}.$$

For piecewise continuous functions on this decomposition, we write

$$v_{\pm}^m = \lim_{s \rightarrow 0_+} v(t_m \pm s), \quad [v]^m = v_+^m - v_-^m.$$

Accordingly, we define the discrete semi-linear form

$$\begin{aligned}A_k(v; \phi) &:= \sum_{m=0}^M \left\{ (\partial_t v, \phi)_m + a_m(v, \phi) + n_m(v, v, \phi) \right\} \\ &\quad + \sum_{m=1}^M ([v]^m, \phi_+^m) + (v_+^0, \phi_+^0),\end{aligned}$$

and the time-discrete Spaces

$$\mathbf{V}_k(I) = \{v_k \in L^2(I; \mathbf{V}), v|_{I_m} \in P_0(I_m), m = 1, \dots, M\}.$$

The time-discrete approximation then seeks a  $v_k \in \mathbf{V}_k(I)$ , such that

$$A_k(v_k; \phi_k) = F(\phi_k) \quad \forall \phi_k \in \mathbf{V}_k(I). \quad (24)$$

We note that this formulation contains the initial condition  $v(0) = v^0$  in the weak sense. The essential feature of the dG(r) schemes are their Galerkin

orthogonality property. Using the fact that the continuous solution  $v$  also satisfies the discrete equation (24), we have

$$A_k(v; \phi_k) - A_k(v_k; \phi_k) = 0, \quad \phi_k \in \mathbf{V}(I). \quad (25)$$

In order to estimate the error  $e := v - v_k$ , we may employ a “parabolic” duality argument. The adjoint of the Fréchet derivative of the governing semi-linear form taken at the solution  $v$  is given by

$$L^*(v; \phi, z) = (\phi, -\partial_t z)_I + a_I(\phi, z) + n_I(v, \phi, z) + n_I(\phi, v, z) - (\phi(T), z(T)).$$

Then, we introduce the “dual solution”  $z \in \mathbf{V}(I)$  as solution of the space-time “dual problem”

$$L^*(v; \phi, z) = (e(T), \phi(T)) \quad \forall \phi \in \mathbf{V}(I). \quad (26)$$

Taking now  $\phi = e$  in (26), we obtain the error representation

$$\begin{aligned} \|e(T)\| &= L^*(v; e, z) \\ &= (e, -\partial_t z)_I + a_I(e, z) + n_I(v, e, z) + n_I(e, v, z). \end{aligned}$$

Using the Galerkin orthogonality (25) and interpolation estimates on each subinterval  $I_m$ , we conclude the estimate

$$\begin{aligned} \|e(T)\| &= \sum_{m=0}^M \left\{ (f, z - z_k)_m - ([v]^m, z_+^m - z_{k,+}^m) \right\} \\ &\leq c_S \left\{ \max_I \|[v]^m\| + \max_I \|kf\| \right\}, \end{aligned}$$

with the “stability constant”  $c_S = c_S(\nu, T, v)$  given by

$$|\log(k_M)|^{-1/2} \left\{ \|z\|_I + |\log(k_M)|^{-1/2} \int_0^{T-k_M} \|\partial_t z\| dt \right\} \leq c_S \|z(T)\|. \quad (27)$$

The result is the “final time” a priori error estimate

$$\|e(T)\| \leq c L_k c_S \max_I \|k \partial_t v\|, \quad (28)$$

where  $L_k := \max_I (1 + \log(k))^{1/2}$ .

**Conclusion:** The foregoing discussion shows that proving a priori error estimates for numerical approximations is closely connected with the study of stability properties of linearizations of the Navier-Stokes equations, i.e. with hydrodynamic stability; for more details see [55]. The goal is to derive a priori estimates of the type (18), (20) and (27) with quantified constants  $c_S = c_S(\nu, T, \text{data})$ . The dependence of these constants on the Reynolds number may be *linear* in “good” cases or may deteriorate to *exponential* in “bad” cases. Such exponentially unstable flows are not computable over relevant intervals of time. The same question will also be crucial for the derivation of *a posteriori* error estimates discussed in Section 7, below.

## 6.5 The problem of stability constants (A critical review of hydrodynamic stability theory)

Most of the traditional stability theory for fluid flow is of qualitative nature being based on eigenvalue criteria through a linearized stability argument. The theoretical results are in good agreement with experiments concerning the *critical* Reynolds number at which the first bifurcation occurs; well-studied examples are the Bénard problem and the Taylor-Couette problems. But they do not fit with experiments for the other fundamental case of parallel flow. There are several paradoxes observed:

- Poiseuille flow (between two parallel plates) is predicted to turn turbulent at  $Re \sim 5772$  through  $2d$  Tollmien-Schlichting waves, while experiments show instability with essential  $3d$  features somewhere in the range  $Re \sim 1000 - 10000$  depending on the experimental setup.
- Couette flow (parallel shear flow) is supposed to be stable for all  $Re > 0$ , but experiments show instability for  $Re \sim 300 - 1500$ .

This failure of theory was blamed on the deficiency of *linearized* stability theory being valid only for *small* perturbations. However, *linearized stability theory* is okay, but was only wrongly interpreted. In dynamic systems governed by *non-normal* matrices, one has to look at the size of the total amplification factors for the initial perturbation and not only at the sign of the eigenvalue's real parts. It is observed that, for example in Poiseuille flow there occurs amplification by a factor of  $10^4$  for  $Re \geq 549$ . Some of the relevant references on this subject are Landahl [59] and Trefethen et al. [92], to mention only a few. More references can be found in [54, 55] where this new concept in hydrodynamic stability theory is discussed in view of numerical approximation. In fact, the question of quantitative hydrodynamic instability and that of numerical computability of laminar flows are closely related. In transition to turbulence one seeks to establish *lower bounds* on the growth of perturbations in order to understand how a laminar flow may develop into a turbulent flow. In error control in CFD for laminar flows one seeks *upper bounds* of the growth of perturbations related to discretizations of the Navier-Stokes equations.

We want to illustrate the phenomenon of error amplification by two simple examples taken from [55].

a) *An ODE model:*

At first, we consider the simple ODE system

$$\dot{w}_1 + \nu w_1 + w_2 = 0, \quad (29)$$

$$\dot{w}_2 + \nu w_2 = 0. \quad (30)$$

Here,  $\nu w_i$  stand for the diffusion terms and  $w_2$  in the first equation for the coupling in the transport term of the linearized perturbation equation of the

Navier-Stokes equations. The corresponding coefficient matrix

$$A = \begin{pmatrix} \nu & 1 \\ 0 & \nu \end{pmatrix}$$

is non-normal; the only eigenvalue  $\lambda = \nu$  has algebraic multiplicity two. This is just the situation described above. For this linear system the solution corresponding to the initial values  $w(0) = w^0$  is given by

$$w_1(t) = e^{-\nu t} w_1^0 - t e^{-\nu t} w_2^0, \quad w_2(t) = e^{-\nu t} w_2^0.$$

We see the exponential decay of the second component and the linear growth over the interval  $[0, \nu^{-1}]$  to size  $w_1(\nu^{-1}) = \nu^{-1} e^{-1} w_2^0$  of the first component before the exponential decay sets in. The component  $w_2$  acts like a catalyst in the first equation. Although exponentially decaying it first causes  $w_1$  to grow; the later exponential decay is irrelevant when by the growth to size  $\nu^{-1} w_1^0$  the linearization is no longer valid.

*b) A simple flow model:*

Next, we consider a very simple configuration: the flow in an infinite pipe  $\Omega = \mathbb{R} \times \omega$  extending in the  $x_1$ -axis with cross section  $\omega$  in the  $(x_2, x_3)$ -plane. The flow is driven by a volume force  $f = (f_1(x_2, x_3, t), 0, 0)^T$  (gravitation) in  $x_1$ -direction. The solution is supposed to have the form (like  $x_1$ -independent Poiseuille flow):

$$v = (v_1(x_2, x_3, t), 0, 0)^T, \quad p = p(x, t).$$

Then the Navier-Stokes equations take the form

$$\partial_t v_1 - \nu \Delta v_1 + \partial_1 p = f_1 \quad \text{in } \omega, \quad v_1|_{\partial\omega} = 0.$$

The corresponding linearized perturbation equation is

$$\begin{aligned} \partial_t w_1 - \nu \Delta w_1 + v_1 \partial_1 w_1 + \partial_2 v_1 w_2 + \partial_3 v_1 w_3 + \partial_1 q &= 0, \\ \partial_t w_2 - \nu \Delta w_2 + v_1 \partial_1 w_2 &+ \partial_2 q = 0, \\ \partial_t w_3 - \nu \Delta w_3 + v_1 \partial_1 w_3 &+ \partial_3 q = 0, \end{aligned}$$

with the incompressibility condition  $\partial_2 w_2 + \partial_3 w_3 = 0$ , and the initial and boundary conditions  $w|_{t=0} = w^0$  and  $w|_{\partial\omega} = 0$ . Even this simple problem is still too complex for an explicit solution. Therefore, we simplify it further by assuming that the perturbed solution  $\{w, q\}$  is also independent of  $x_1$ . This corresponds to looking at a fluid in a long vertical tube under gravity or in a long rotating tube with varying speed of rotation. Under this assumption the perturbation equation reduces to

$$\begin{aligned} \partial_t w_1 - \nu \Delta w_1 + \partial_2 v_1 w_2 + \partial_3 v_1 w_3 &= 0, \\ \partial_t w_2 - \nu \Delta w_2 &+ \partial_2 q = 0, \\ \partial_t w_3 - \nu \Delta w_3 &+ \partial_3 q = 0. \end{aligned}$$

In this system the equations for the components  $\bar{w} := \{w_2, w_3\}$  together with the constraint  $\partial_2 w_2 + \partial_3 w_3 = 0$  form a two-dimensional Stokes problem which can be solved independently of the first equation. Hence, we are in a similar situation as in the above ODE example. For the Stokes subsystem we have the standard a priori estimate

$$\|\bar{w}(t)\| \leq e^{-\kappa\nu t} \|\bar{w}^0\|, \quad t \geq 0,$$

with  $\kappa = \text{diam}(\omega)$ . The first equation does not contain the pressure. Using the result for  $\bar{w}$  we obtain for the first component  $w_1$  the bound

$$\|w_1(t)\| \leq ce^{-\kappa\nu t} \{t\|w_1^0\| + \|\bar{w}^0\|\}, \quad t \geq 0. \quad (31)$$

Hence, we see that for this model problem, one can show that the error constant in the a priori error estimate (1) grows at most linearly with the Reynolds number:

$$C(\nu, T, \text{data}) \sim \max\{T, Re\}.$$

It is an open question whether this linear dependence on  $Re$  is generic for a larger class of flow problems. Numerical experiments for the lid-driven cavity flow show such a dependence.

**Open Problem 6.4:** *Prove a posteriori stability estimates like (31) for more practical problems (e.g. Poiseuille flow) possibly with respect to different norms. Is there any indication that linear growth in time of perturbations may be generic to the Navier-Stokes equations?*

## 7 Error control and mesh adaptation

This section is devoted to concepts of error estimation and mesh optimization. The goal is to develop techniques for reliable estimation of the discretization error in quantities of physical interest as well as economical mesh adaptation. The use of a finite element Galerkin discretization provides the appropriate framework for a mathematically rigorous error analysis. On the basis of computable a posteriori error bounds the mesh is locally refined within a feed-back process yielding economical mesh-size distributions for prescribed error tolerance or maximum number of cells. On the resulting sequence of refined meshes the discrete problems are solved by multi-level techniques.

The general concept of residual-based error control for finite element methods is described in the survey article by Eriksson/Estep/Hansbo/Johnson [26]; this technique has then been further developed for various situations in [12, 14]. The application to incompressible flows is extensively discussed in Becker [7, 9]. Extensions to compressible flow including chemical reactions are given in Braack [17]; see also [18]. A survey of applications of this approach to a variety of other problems can be found in [75].

### 7.1 Principles of error estimation

The discretization error in a cell  $K$  splits into two components, the *locally* produced error (truncation error) and the *transported* error (pollution error)

$$e_K^{tot} = e_K^{loc} + e_K^{trans}. \quad (1)$$

The effect of the cell residual  $\rho_K$  on the local error  $e_{K'}$ , at another cell  $K'$ , is governed by the Green function of the continuous problem. This is the general philosophy underlying our approach to error control.

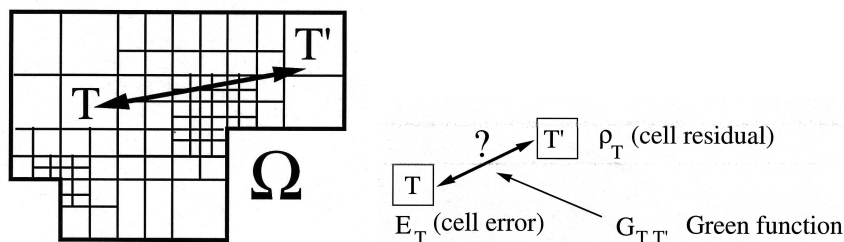


Figure 28: Scheme of error propagation

(I) *A priori error analysis*: The classical *a priori* error estimation aims at estimating the error to be expected in a computation which is still to be done. These bounds are expressed in terms of powers of a mesh size  $h$  and involve constants which depend on the (unknown) exact solution. In this way, only asymptotic (as  $h \rightarrow 0$ ) information about the error behavior is provided but

no quantitatively useful error bound. In particular, no criterion for local mesh adaptation is obtained.

(II) *A posteriori error analysis:* The *a posteriori* error analysis generates error estimates in the course of the computation. Accordingly, these bounds are in terms of computable local residuals of the approximate solution and do not require information about the exact solution. However, a posteriori error analysis usually does not provide a priori information about the convergence of the discretization process as  $h \rightarrow 0$ .

We illustrate the basic principles underlying error estimation by considering perturbations of linear algebraic systems. Let  $A, \tilde{A} \in \mathbb{R}^{n \times n}$ ,  $b, \tilde{b} \in \mathbb{R}^n$  be given and solve

$$Ax = b, \quad \tilde{A}\tilde{x} = \tilde{b} \quad (\text{perturbed problem}). \quad (2)$$

For estimating the error  $e := x - \tilde{x}$ , there are several approaches. The *a priori* method uses the “truncation error”  $\tau := \tilde{A}x - \tilde{b} = \tilde{A}(x - \tilde{x})$ ,

$$e = \tilde{A}^{-1}\tau \quad \Rightarrow \quad \|e\| \leq \tilde{c}_S \|\tau\|, \quad (3)$$

with the “discrete” stability constant  $\tilde{c}_S := \|\tilde{A}^{-1}\|$ . The *a posteriori* method uses the “residual”  $\rho := b - A\tilde{x} = A(x - \tilde{x})$ ,

$$e = A^{-1}\rho \quad \Rightarrow \quad \|e\| \leq c_S \|\rho\|, \quad (4)$$

with the “continuous” stability constant  $c_S := \|A^{-1}\|$ . Alternatively, we may use the solution  $z$  of the “dual problem”  $A^*z = \|e\|^{-1}e$ , to obtain

$$\|e\| = (e, A^*z) = (b - A\tilde{x}, z) = (\rho, z) \leq \|\rho\| \|z\| \leq c_S^* \|\rho\|, \quad (5)$$

with the “dual” stability constant  $c_S^* := \|A^{*-1}\|$ . Of course, this approach does not yield a new result in estimating the error in the  $l_2$ -norm. But it shows the way to bound other error quantities as for example single components  $|e_i|$ .

An analogous argument can also be applied in the case of nonlinear equations. Let  $F, \tilde{F} : \mathbb{R}^n \rightarrow \mathbb{R}^n$  be (differentiable) vector functions and solve

$$F(x) = b, \quad \tilde{F}(\tilde{x}) = \tilde{b} \quad (\text{perturbed problem}). \quad (6)$$

Then, the residual  $\rho := b - F(\tilde{x})$  satisfies

$$\rho = F(x) - F(\tilde{x}) = \left( \int_0^1 F'(\tilde{x} + se) ds \right) e =: L(x, \tilde{x})e, \quad (7)$$

with the Jacobian  $F'$ . The term in parentheses defines a linear operator  $L(x, \tilde{x}) : \mathbb{R}^n \rightarrow \mathbb{R}^n$  which depends on the (unknown) solution  $x$ . It follows that  $\|e\| \leq c_S \|\rho\|$ , with the (nonlinear) stability constant  $c_S := \|L(x, \tilde{x})^{-1}\|$ . Below, we will use this duality technique for generating a posteriori error estimates in Galerkin finite element methods for differential equations.



### 7.1.1 A DIFFUSION MODEL PROBLEM

For illustrating our concept, we start with the (scalar) model diffusion problem

$$-\Delta u = f \text{ in } \Omega, \quad u = 0 \text{ on } \partial\Omega, \quad (8)$$

posed on a polygonal domain  $\Omega \subset \mathbb{R}^2$ . In its variational formulation one seeks  $u \in V := H_0^1(\Omega)$  satisfying

$$(\nabla u, \nabla \phi) = (f, \phi) \quad \forall \phi \in V. \quad (9)$$

We consider a finite element approximation using piecewise (isoparametric) bilinear shape functions (see Section 3). The corresponding finite element spaces  $V_h \subset V$  are defined on decompositions  $\mathbb{T}_h$  of  $\bar{\Omega}$  into quadrilaterals (“cells”)  $K$  of width  $h_K := \text{diam}(K)$ . We write again  $h := \max_{K \in \mathbb{T}} h_K$  for the maximal *global* mesh width. Simultaneously, the notation  $h = h(x)$  is used for the continuously distributed mesh-size function defined by  $h|_K = h_K$ . For ease of mesh refinement and coarsening we allow “hanging nodes”, but at most one per edge. The shape of the corresponding modified basis function is shown in Figure 29.

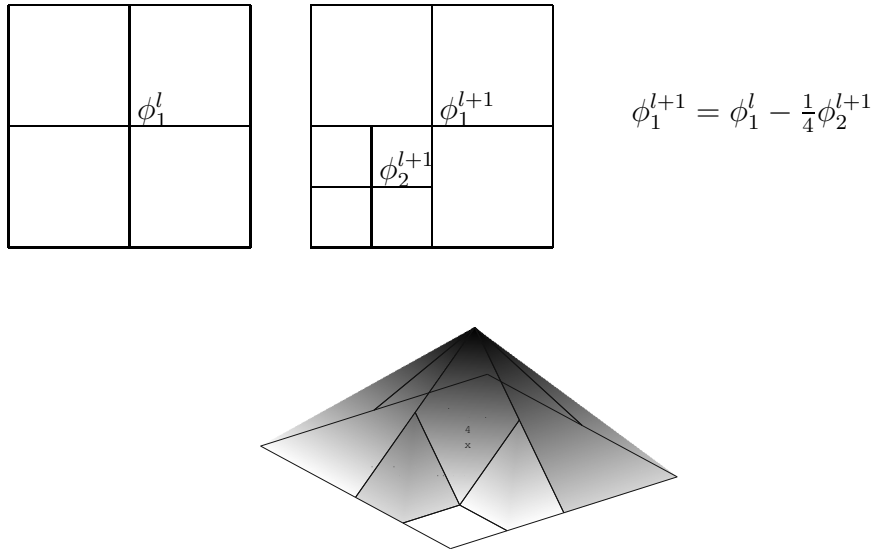


Figure 29:  $Q_1$  nodal basis function on a patch of cells with a hanging node

The discrete problem determines  $u_h \in V_h$  by

$$(\nabla u_h, \nabla \phi_h) = (f, \phi_h) \quad \forall \phi_h \in V_h. \quad (10)$$

We recall the “Galerkin orthogonality” of the error  $e := u - u_h$ ,

$$(\nabla e, \nabla \phi_h) = 0, \quad \phi_h \in V_h. \quad (11)$$

We seek to derive a posteriori error estimates. Let  $J(\cdot)$  be an arbitrary “error functional” defined on  $V$  and  $z \in V$  the solution of the corresponding dual problem

$$(\nabla\phi, \nabla z) = J(\phi) \quad \forall \phi \in V. \quad (12)$$

Setting  $\phi = e$  in (12) results in the error representation

$$\begin{aligned} J(e) &= (\nabla e, \nabla z) = (\nabla e, \nabla(z - I_h z)) \\ &= \sum_{K \in \mathbb{T}} \left\{ (-\Delta u + \Delta u_h, z - I_h z)_K - (\partial_n u_h, z - I_h z)_{\partial K} \right\} \\ &= \sum_{K \in \mathbb{T}} \left\{ (f + \Delta u_h, z - I_h z)_K - \frac{1}{2}(n \cdot [\nabla u_h], z - I_h z)_{\partial K} \right\}, \end{aligned} \quad (13)$$

where  $[\nabla u_h]$  is the jump of  $\nabla u_h$  across the interelement boundary. In the second equation, we have used galerkin orthogonality. This gives us the a posteriori error estimate

$$|J(e)| \leq \eta(u_h) := \sum_{K \in \mathbb{T}_h} h_K^4 \left\{ \rho_K(u_h) \omega_K(z) + \rho_{\partial K}(u_h) \omega_{\partial K}(z) \right\}, \quad (14)$$

with the cell residuals

$$\rho_K(u_h) := h_K^{-1} \|f + \Delta u_h\|_K, \quad \rho_{\partial K}(u_h) := h_K^{-3/2} \|n \cdot [\nabla u_h]\|_{\partial K},$$

and the weights

$$\omega_K(z) := h_K^{-3} \|z - I_h z\|_K, \quad \omega_{\partial K}(z) := \frac{1}{2} h_K^{-5/2} \|z - I_h z\|_{\partial K}.$$

These quantities are normalized, such that they can be expected to approach certain mesh-independent limits as  $h \rightarrow 0$ . The interpretation of the relation (14) is that the weights  $\omega_K(z)$  describe the dependence of  $J(e)$  on variations of the cell residuals  $\rho_K(u_h)$ ,

$$\frac{\partial J(e)}{\partial \rho_K} \approx h_K^4 \omega_K(z) \approx h_K^4 \max_K |\nabla^2 z|.$$

We remark that in a finite difference discretization of the model problem (8) the corresponding “influence factors” behave like  $\omega_K(z) \approx h_K^2 \max_K |z|$ .

In practice the weights  $\omega_K(z)$  have to be determined computationally. Let  $z_h \in V_h$  be the finite element approximation of  $z$ ,

$$(\nabla\phi_h, \nabla z_h) = J(\phi_h) \quad \forall \phi_h \in V_h. \quad (15)$$

We can estimate

$$\omega_K(z) \leq c_I h_K^{-1} \|\nabla^2 z\|_K \approx c_I \max_K |\nabla_h^2 z_h|, \quad (16)$$

where  $\nabla_h^2 z_h$  is a suitable difference quotient approximating  $\nabla^2 z$ . The interpolation constant is usually in the range  $c_I \approx 0.1 - 1$  and can be determined by calibration. Alternatively, we may construct from  $z_h \in V_h$  a patchwise biquadratic interpolation  $I_h^{(2)} z_h$  and replace  $z - I_h z$  in the weight  $\omega_K(z)$  by  $I_h^{(2)} z_h - z_h$ . This gives an approximation to  $\omega_K(z)$  which is free of interpolation constants.

One may try to further improve the quality of the error estimate by solving local defect equations, either Dirichlet problems (à la Babuska/Miller) or Neumann problems (à la Bank/Weiser); see Backes [4]. References for these approaches are Verfürth [102] and Ainsworth/Oden [1]. Comparison with simpler mesh adaptation techniques, e.g. refinement criteria based on difference quotients of the computed solution, local gradient recovery “ZZ technique” (à la Zienkiewicz/Zhu [106]), or other local “ad hoc” criteria have been reported in Braack [17] and in [75].

By the same type of argument, one can also derive the traditional global error estimates in the energy and the  $L^2$  norm.

(i) *Energy-norm error bound:* Using the functional

$$J(\phi) := \|\nabla e\|^{-1}(\nabla e, \nabla \phi)$$

in the dual problem, we obtain the estimate

$$\|\nabla e\| \leq \sum_{K \in \mathbb{T}} h_K^4 \rho_K(u_h) \omega_K(z) \leq c_I \sum_{K \in \mathbb{T}} h_K^2 \rho_K(u_h) \|\nabla z\|_{\tilde{K}},$$

where  $\tilde{K}$  is the union of all cells neighboring  $K$ . In view of the a priori bound  $\|\nabla z\| \leq c_S = 1$ , this implies the a posteriori error estimate

$$\|\nabla e\| \leq \eta_E(u_h) := c_I \left( \sum_{K \in \mathbb{T}} h_K^4 \rho_K(u_h)^2 \right)^{1/2}. \quad (17)$$

(ii)  *$L^2$ -norm error bounds:* Using the functional

$$J(\phi) := \|e\|^{-1}(e, \phi)$$

in the dual problem, we obtain the estimate

$$\|e\| \leq \sum_{K \in \mathbb{T}} h_K^4 \rho_K(u_h) \omega_K(z) \leq c_I \sum_{K \in \mathbb{T}} h_K^3 \rho_K(u_h) \|\nabla^2 z\|_K.$$

In view of the a priori bound  $\|\nabla^2 z\| \leq c_S$  ( $c_S = 1$  if  $\Omega$  is convex), this implies the a posteriori error bound

$$\|e\| \leq \eta_{L^2}(u_h) := c_I c_S \left( \sum_{K \in \mathbb{T}} h_K^6 \rho_K(u_h)^2 \right)^{1/2}. \quad (18)$$

### 7.1.2 A TRANSPORT MODEL PROBLEM

As a simple model, we consider the scalar transport equation

$$\beta \cdot \nabla u = f, \quad (19)$$

on a domain  $\Omega \subset \mathbb{R}^2$  with inflow boundary condition  $u = g$  along the “inflow boundary”  $\partial\Omega_- = \{x \in \partial\Omega, n \cdot \beta < 0\}$ . Accordingly,  $\partial\Omega_+ = \partial\Omega \setminus \partial\Omega_-$  is the “outflow boundary”. The transport vector  $\beta$  is assumed as constant for simplicity; therefore, the natural solution space is

$$V := \{v \in L^2(\Omega), \beta \cdot \nabla v \in L^2(\Omega)\}.$$

This problem is discretized using the Galerkin finite element method with streamline diffusion stabilization as described above. On quadrilateral meshes  $\mathbb{T}_h$ , we define again subspaces  $V_h = \{v \in H^1(\Omega), v|_K \in \tilde{Q}_1(K), K \in \mathbb{T}_h\}$ , where  $\tilde{Q}_1$  is the space of “isoparametric” bilinear functions on cell  $K$ . The discrete solution  $u_h \in V_h$  is defined by

$$(\beta \cdot \nabla u_h - f, \phi + \delta \beta \cdot \nabla \phi) + (n \cdot \beta (g - u_h), \phi)_{\partial\Omega_-} = 0 \quad \forall \phi \in V_h, \quad (20)$$

where the stabilization parameter is determined locally by  $\delta_K = h_K$ . In this formulation the inflow boundary condition is imposed in the weak sense. This facilitates the use of a duality argument in generating a posteriori error estimates. Let  $J(\cdot)$  be a given functional with respect to which the error  $e = u - u_h$  is to be controlled. Following our general approach, we consider the corresponding dual problem

$$(\beta \cdot \nabla \phi, z + \delta \beta \cdot \nabla z) - (n \cdot \beta \phi, z)_{\partial\Omega_-} = J(\phi) \quad \forall \phi \in V, \quad (21)$$

which is a transport problem with transport in the negative  $\beta$ -direction. We note that the stabilized bilinear form  $A_h(\cdot, \cdot)$  is used in the duality argument, in order to achieve an optimal treatment of the stabilization terms; for a detailed discussion of this point see [75] and [47]. The error representation reads

$$J(e) = (\beta \cdot \nabla e, z - z_h + \delta \beta \cdot \nabla (z - z_h)) - (n \cdot \beta e, z - z_h)_{\partial\Omega_-},$$

for arbitrary  $z_h \in V_h$ . This results in the a posteriori error estimate

$$|J(e)| \leq \eta(u_h) := c_I \sum_{K \in \mathbb{T}} h_K^4 \left\{ \rho_K(u_h) \omega_K(z) + \rho_{\partial K}(u_h) \omega_{\partial K}(z) \right\}, \quad (22)$$

with the cell residuals

$$\rho_K(u_h) := h_K^{-1} \|f - \beta \cdot \nabla u_h\|_K, \quad \rho_{\partial K}(u_h) := h_K^{-3/2} \|n \cdot \beta (u_h - g)\|_{\partial K \cap \partial\Omega_-},$$

and cell weights (setting  $\xi := z - z_h$ )

$$\omega_K(z) := h_K^{-3} \left\{ \|\xi\|_K + \delta_K \|\beta \cdot \nabla \xi\|_K \right\}, \quad \omega_{\partial K}(z) := h_K^{-5/2} \|\xi\|_{\partial K \cap \partial\Omega_-}.$$

We note that this a posteriori error bound explicitly contains the mesh size  $h_K$  and the stabilization parameter  $\delta_K$  as well. This gives us the possibility to simultaneously adapt both parameters, which may be particularly advantageous in capturing sharp layers in the solution.

We want to illustrate the features of the error estimate (22) by a simple thought experiment. Let  $\Omega = (0, 1) \times (0, 1)$  and  $f = 0$ . We take the functional

$$J(u) := (1, n \cdot \beta u)_{\partial\Omega_+}.$$

The corresponding dual solution is  $z \equiv 1$ , so that  $J(e) = 0$ . This implies

$$(1, n \cdot \beta u_h)_{\partial\Omega_+} = (1, n \cdot \beta u)_{\partial\Omega_+} = -(1, n \cdot \beta g)_{\partial\Omega_-}.$$

recovering the well-known global conservation property of the scheme.

### 7.1.3 EVALUATION OF THE ERROR ESTIMATES

To evaluate the error estimates (14) or (22), one may solve the corresponding perturbed dual problem numerically by the same method as used in computing  $u_h$ , yielding an approximation  $z_h \in V_h$  to the exact dual solution  $z$ . However, the use of the same meshes for computing primal and dual solution is by no means obligatory. In fact, in the case of dominant transport it may be advisable to compute the dual solution on a different mesh; see [47] for examples. Then, the weights  $\omega_K$  can be determined numerically in different ways:

1. We may take  $z_h = I_h z \in V_h$  as the nodal interpolation of  $z$  and use the local interpolation properties of finite elements to obtain

$$\omega_K = h_K^{-3} \|z - I_h z\|_K \leq c_I h_K^{-1} \|\nabla^2 z\|_K,$$

with an interpolation constant  $c_I \approx 0.1 - 1$ . Here,  $\nabla^2 z$  is the tensor of second derivatives of  $z$ . Then, approximation by second-order difference quotients of the computed discrete dual solution  $z_h \in V_h$  yields

$$\omega_K \approx c_I |\nabla_h^2 z_h(x_K)|, \quad (23)$$

$x_K$  being the center point of  $K$ .

2. Computation of a discrete dual solution  $z_{h'} \in V_{h'}$  in a richer space  $V_{h'} \supset V_h$  (e.g., on a finer mesh or by higher-order elements) and setting

$$\omega_K \approx h_K^{-3} \|z_{h'} - I_h z_{h'}\|_K, \quad (24)$$

where  $I_h z_{h'} \in V_h$  denotes the generic nodal interpolation.

3. Interpolation of the discrete dual solution  $z_h \in V_h$  by higher order polynomials on certain cell-patches, e.g., biquadratic interpolation  $I_h^{(2)} z_h$ :

$$\omega_K \approx h_K^{-3} \|I_h^{(2)} z_h - z_h\|_K. \quad (25)$$

Analogous approximations can be used for the weights  $\omega_{\partial K}$ . Option (2) is quite expensive and rarely used. Since we normally do not want to spend more time in evaluating the error estimate than for solving the primal problem, we recommend option (1) or (3). Notice that option (3) does not involve an interpolation constant which needs to be specified. The computational results reported in [14] indicate that the use of biquadratic interpolation on patches of four quadrilaterals is more accurate than using the finite difference approximation (23).

## 7.2 Strategies for mesh adaptation

We use the notation introduced above:  $u$  is the solution of the variational problem posed on a 2-dimensional domain  $\Omega$ ,  $u_h$  is its piecewise linear (or bilinear) finite element approximation. Further,  $e = u - u_h$  is the discretization error and  $J(\cdot)$  a linear error functional for measuring  $e$ . We suppose that there is an a posteriori error estimate of the form

$$|J(e)| \leq \eta := \sum_{K \in \mathcal{T}_h} h_K^4 \rho_K(u_h) \omega_K(z), \quad (26)$$

with the cell residuals  $\rho_K(u_h)$  and weights  $\omega_K(z)$ . Accordingly, we define the local “error indicators”

$$\eta_K := h_K^4 \rho_K(u_h) \omega_K(z).$$

The mesh design strategies are oriented towards a prescribed tolerance  $TOL$  for the error quantity  $J(e)$  and the number of mesh cells  $N$  which measures the complexity of the computational model. Usually the admissible complexity is constrained by some maximum value  $N_{\max}$ .

There are various strategies for organizing a mesh adaptation process on the basis of the a posteriori error estimate (26).

- *Error balancing strategy:* Cycle through the mesh and equilibrate the local error indicators,

$$\eta_K \approx \frac{TOL}{N} \quad \Rightarrow \quad \eta \approx TOL. \quad (27)$$

This process requires iteration with respect to the number of cells  $N$ .

- *Fixed fraction strategy:* Order cells according to the size of  $\eta_K$  and refine a certain percentage (say 30%) of cells with largest  $\eta_K$  (or those which make up 30% of the estimate value  $\eta$ ) and coarsen those cells with smallest  $\eta_K$ . By this strategy, we may achieve a prescribed rate of increase of  $N$  (or keep it constant as may be desirable in nonstationary computations).

- *Mesh optimization strategy:* Use the representation

$$\eta := \sum_{K \in \mathcal{T}_h} h_K^4 \rho_K(u_h) \omega_K(z) \approx \int_{\Omega} h(x)^2 A(x) dx \quad (28)$$

for generating a formula for an optimal mesh-size distribution  $h_{opt}(x)$ .

We want to discuss the strategy for deriving an optimal mesh-size distribution in more detail. As a side-product, we will also obtain the justification of the error equilibration strategy. Let  $N_{max}$  and  $TOL$  be prescribed. We assume that for  $TOL \rightarrow 0$ , the cell residuals and the weights approach certain limits,

$$\begin{aligned} \rho_K(u_h) &\approx h_K^{-3/2} \|n \cdot [\nabla u_h]\|_{\partial K} \rightarrow \rho(x_K) \approx |D^2 u(x_K)|, \\ \omega_K(z) &\approx h_K^{-5/2} \|z - I_h z\|_{\partial K} \rightarrow \omega(x_K) \approx |D^2 z(x_K)|. \end{aligned}$$

These properties can be proven on uniformly refined meshes by exploiting super-convergence effects, but still need theoretical justification on locally refined meshes. This suggests to assume that

$$\eta \approx \tilde{\eta} := \int_{\Omega} h(x)^2 A(x) dx, \quad N = \sum_{K \in \mathcal{T}_h} h_K^2 h_K^{-2} \approx \int_{\Omega} h(x)^{-2} dx, \quad (29)$$

with the weighting function  $A(x) = \rho(x)\omega(x)$ . Now, let us consider the mesh optimization problem

$$\eta \rightarrow \min!, \quad N \leq N_{max}.$$

Applying the usual Lagrange approach yields the necessary optimality conditions

$$\frac{d}{dt} \left[ \int_{\Omega} (h + t\phi)^2 A dx + (\lambda + t\mu) \left( (h + t\mu)^{-2} dx - N_{max} \right) \right]_{t=0} = 0,$$

for any variations  $\phi$  and  $\mu$ . From this, we infer that

$$2h(x)A(x) - 2\lambda h(x)^{-3} = 0, \quad \int_{\Omega} h(x)^{-2} dx - N_{max} = 0.$$

Hence, we obtain

$$h(x) = (\lambda^{-1} A(x))^{-1/4} \Rightarrow \tilde{\eta} = h^4 A \equiv \lambda^{-1},$$

and

$$\lambda^{-1/2} \int_{\Omega} A(x)^{1/2} dx = N_{max}, \quad W := \int_{\Omega} A(x)^{1/2} dx.$$

This gives us a formula for the ‘‘optimal’’ mesh-size distribution:

$$\lambda = \left( \frac{W}{N_{max}} \right)^2 \Rightarrow h_{opt}(x) = \left( \frac{W}{N_{max}} \right)^{1/2} A(x)^{-1/4}. \quad (30)$$

In an analogous way, we can also treat the adjoint optimization problem  $N \rightarrow \min!$ ,  $\eta \leq TOL$ . We note that even for a rather “singular” error functional  $J(\cdot)$  the quantity  $W$  is bounded, e.g.,

$$J(e) = \nabla e(0) \quad \Rightarrow \quad A(x) \approx |x|^{-3} \quad \Rightarrow \quad W = \int_{\Omega} |x|^{-3/2} dx < \infty.$$

**Open Problem 7.1:** *Make the “mesh optimization strategy” rigorous, i.e., prove the proposed convergence of cell weights and residuals under (local) mesh refinement. This could be accomplished by proving that for piecewise linear or  $d$ -linear approximation, there holds*

$$\overline{\lim}_{h \rightarrow 0} \left\{ \max_{\Omega} |\nabla_h^2 u_h| \right\} \leq c(u),$$

where  $\nabla_h^2 u_h$  is a suitable second-order difference quotient.

### 7.2.1 COMPUTATIONAL TESTS

(I) *The diffusion model problem:* We begin with the model diffusion problem (8) posed on the rectangular domain  $\Omega = (-1, 1) \times (-1, 3)$  with slit at  $(0, 0)$ . In the presence of a reentrant corner, here a slit, with angle  $\omega = 2\pi$ , the solution involves a “corner singularity”. It can be written in the form  $u = \psi r^{1/2} + \tilde{u}$ , with  $r$  being the distance to the corner point and  $\tilde{u} \in H^2(\Omega)$ . We want to illustrate how the singularity introduced by the weights interacts with the pollution effect caused by the slit singularity. Let the goal be the accurate computation of the a derivative value  $J(u) = \partial_1 u(P)$  at the point  $P = (0.75, 2.25)$ . In this case the dual solution  $z$  behaves like

$$|\nabla^2 z(x)| \approx d(x)^{-3} + r(x)^{-3/2},$$

where  $d(x)$  and  $r(x)$  are the distance functions with respect to the points  $P$  and  $(0, 0)$ , respectively. Notice that in this case, the dual solution does not exist in the sense of  $H_0^1(\Omega)$ , such that for practical use, we have to regularize the functional  $J(u) = \partial_1 u(P)$  appropriately. It follows that

$$|\partial_1 e(P)| \approx c_I \sum_{K \in \mathbb{T}_h} h_K^4 \rho_K(u_h) \left\{ d_K^{-3} + r_K^{-3/2} \right\}. \quad (31)$$

Equilibrating the local error indicators yields

$$\eta_K \approx \frac{h_K^4}{d_K^3} \approx \frac{TOL}{N} \quad \Rightarrow \quad h_K^2 \approx d_K^{3/2} \left( \frac{TOL}{N} \right)^{1/2},$$

and, consequently,

$$N = \sum_{K \in \mathbb{T}_h} h_K^2 h_K^{-2} = \left( \frac{N}{TOL} \right)^{1/2} \sum_{K \in \mathbb{T}_h} h_K^2 d_K^{-3/2} \approx \left( \frac{N}{TOL} \right)^{1/2}.$$



This implies that  $N_{opt} \approx TOL^{-1}$  which is better than what could be achieved on a uniformly refined mesh. In fact, the *global* energy-error estimate leads to a mesh efficiency like  $J(e) \sim N^{-1/2}$ , i.e.,  $N_{opt} \approx TOL^{-2}$ . This predicted asymptotic behavior is well confirmed by the results of our computational test shown in Figures 30 and 31 (for more details, we refer to [14]).

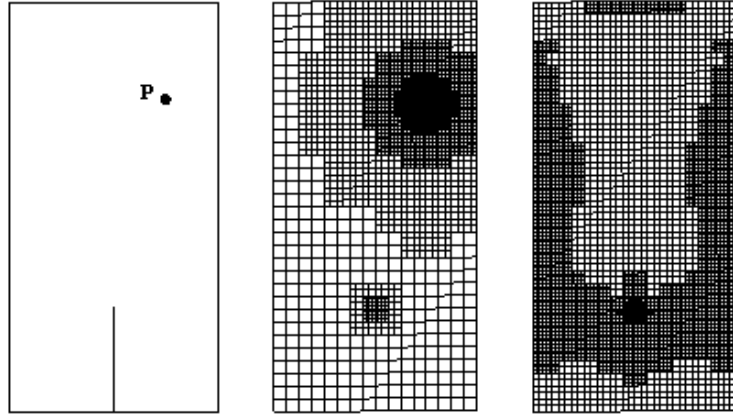


Figure 30: Refined meshes with about 5,000 cells for computing  $\partial_1 u(P)$  using the weighted error estimate  $\eta_{weight}$  (middle) and the energy error estimate  $\eta_E$  (right); from Backes [4].

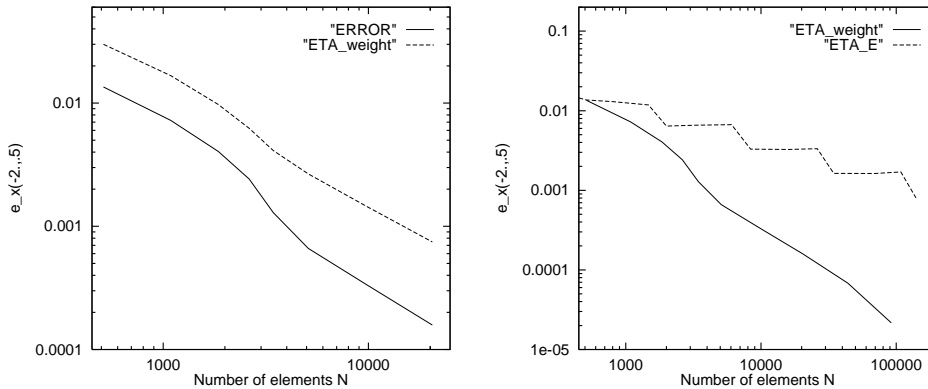


Figure 31: Test results on the slit domain obtained by the weighted error estimator  $\eta_{weight}$ : comparison with the true error (left) and comparison of efficiency with that of the energy error estimator  $\eta_E$  (right); from Backes [4].

(II) *The transport model problem:* Next, we consider the model problem (19) on the unit square  $\Omega = (0, 1) \times (0, 1) \subset \mathbb{R}^2$  with the right-hand side  $f \equiv 0$ , the (constant) transport coefficient  $\beta = (1, 0.5)^T$ , and the inflow data

$$g(x, 0) = 0, \quad g(0, y) = 1.$$

The quantity to be computed is part of the outflow as indicated in Figure 32:

$$J(u) := \int_{\Gamma} \beta \cdot nu \, ds.$$

The mesh refinement is organized according to the “fixed fraction strategy” described above. In Table 3, we show results for this test computation. The corresponding meshes and the primal as well as the dual solution are presented in Figure 32. Notice that there is no mesh refinement enforced along the upper line of discontinuity of the dual solution since here the residual of the primal solution is almost zero.

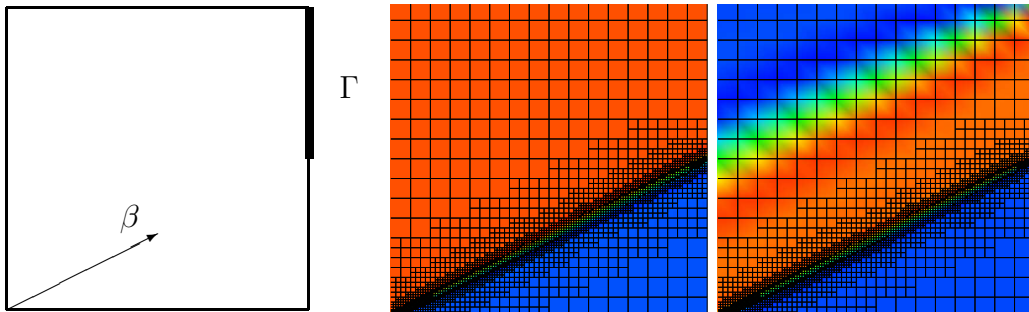


Figure 32: Configuration and grids of the test computation for the model transport problem (19): primal solution (left) and dual solution (right) on an adaptively refined mesh.

Table 3: Convergence results of the test computation for the model transport problem (19).

Level	$N$	$J(e)$	$\eta$	$\eta/J(e)$
0	256	2.01e-2	2.38e-2	1.18
1	310	1.82e-2	1.96e-2	1.08
2	634	1.09e-2	1.21e-2	1.11
3	964	7.02e-3	8.23e-3	1.17
4	1315	6.25e-3	7.88e-3	1.26
5	1540	5.37e-3	6.94e-3	1.29
6	2050	4.21e-3	5.37e-3	1.27
7	2128	4.11e-3	5.21e-3	1.27

### 7.3 A general paradigm for a posteriori error estimation

The approach to residual-based error estimation described above for the linear model problem can be extended to general nonlinear systems. We outline the underlying concept in an abstract setting following the general paradigm introduced by Johnson, et al. [26].

Let  $V$  be a Hilbert space with inner product  $(\cdot, \cdot)$  and corresponding norm  $\|\cdot\|$ ,  $A(\cdot; \cdot)$  a continuous semi-linear form and  $F(\cdot)$  a linear form defined on  $V$ . We seek a solution  $u \in V$  to the abstract variational problem

$$A(u; \phi) = F(\phi) \quad \forall \phi \in V. \quad (32)$$

This problem is approximated by a Galerkin method using a sequence of finite dimensional subspaces  $V_h \subset V$  parameterized by a discretization parameter  $h$ . The discrete problems seek  $u_h \in V_h$  satisfying

$$A(u_h; \phi_h) = F(\phi_h) \quad \forall \phi_h \in V_h. \quad (33)$$

The key feature of this approximation is the ‘‘Galerkin orthogonality’’ which in this nonlinear case reads as

$$A(u; \phi_h) - A(u_h; \phi_h) = 0, \quad \phi_h \in V_h. \quad (34)$$

By elementary calculus, there holds

$$A(u; \phi) - A(u_h; \phi) = \int_0^1 A'(su + (1-s)u_h; e, \phi) ds,$$

with  $A'(v; \cdot, \cdot)$  denoting the tangent form of  $A(\cdot; \cdot)$  at some  $v \in V$ . This leads us to introduce the bilinear form

$$L(u, u_h; \phi, \psi) := \int_0^1 A'(su + (1-s)u_h; \phi, \psi) ds,$$

which depends on the solutions  $u$  as well as  $u_h$ . Then, denoting the error by  $e = u - u_h$ , there holds

$$\begin{aligned} L(u, u_h; e, \phi_h) &= \int_0^1 A'(su + (1-s)u_h; e, \phi_h) ds \\ &= A(u; \phi_h) - A(u_h; \phi_h) = 0, \quad \phi_h \in V_h. \end{aligned}$$

Suppose that the quantity  $J(u)$  has to be computed, where  $J(\cdot)$  is a linear functional defined on  $V$ . For representing the error  $J(e)$ , we use the solution  $z \in V$  of the *dual problem*

$$L(u, u_h; \phi, z) = J(\phi) \quad \forall \phi \in V. \quad (35)$$

Assuming that this problem is solvable and using the Galerkin orthogonality (34), we obtain the error representation

$$J(e) = L(u, u_h; e, z - z_h) = F(z - z_h) - A(u_h; z - z_h), \quad (36)$$

with any approximation  $z_h \in V_h$ . Since the bilinear form  $L(u, u_h; \cdot, \cdot)$  contains the unknown solution  $u$  in its coefficient, the evaluation of (36) requires approximation. The simplest way is to replace  $u$  by  $u_h$ , yielding a perturbed dual problem

$$L(u_h, u_h; \phi, \tilde{z}) = J(\phi) \quad \forall \phi \in V. \quad (37)$$

We remark that the bilinear form  $L(u_h, u_h; \cdot, \cdot)$  used in (37) is identical to the tangent form  $A'(u_h; \cdot, \cdot)$ . Controlling the effect of this perturbation on the accuracy of the resulting error estimate may be a delicate task and depends strongly on the particular problem under consideration. Our own experience with several different types of problems (including the Navier-Stokes equations) indicates that this problem is less critical as long as the continuous solution is stable. The crucial problem is the numerical computation of the perturbed dual solution  $\tilde{z}$  by solving a discretized dual problem

$$L(u_h, u_h; \phi, \tilde{z}_h) = J(\phi) \quad \forall \phi \in V_h. \quad (38)$$

This then results in a practically useful error estimate  $J(e) \approx \tilde{\eta}(u_h)$ .

**Open Problem 7.2:** *Analyze the effect of the error introduced by the linearization steps (37) and (38) on the quality of the a posteriori error bound and design a reliable strategy for controlling this error.*

### 7.3.1 THE NESTED SOLUTION APPROACH

For solving the nonlinear problem (32) by the adaptive Galerkin finite element method (33), we employ the following iterative scheme. Starting from a coarse initial mesh  $\mathbb{T}_0$ , a hierarchy of refined meshes  $\mathbb{T}_i$ ,  $i \geq 1$ , and corresponding finite element spaces  $V_i$  is generated by a nested solution process.

(0) *Initialization*  $i = 0$ : Start on coarse mesh  $\mathbb{T}_0$  with

$$v_0^{(0)} \in V_0.$$

(1) *Defect correction iteration:* For  $i \geq 1$ , start with

$$v_i^{(0)} = v_{i-1} \in V_i.$$

(2) *Iteration step:* Evaluate the defect

$$(d_i^{(j)}, \phi) := F(\phi) - A(v_i^{(j)}; \phi), \quad \phi \in V_i.$$

Choose a suitable approximation  $\tilde{A}'(v_i^{(j)}; \cdot, \cdot)$  to the derivative  $A'(v_i^{(j)}; \cdot, \cdot)$  (with good stability and solubility properties) and solve the correction equation

$$\delta v_i^{(j)} \in V_i: \quad \tilde{A}'(v_i^{(j)}; \delta v_i^{(j)}, \phi) = (d_i^{(j)}, \phi) \quad \forall \phi \in V_i.$$

For this, Krylov-space or multigrid iterations are employed using the hierarchy of already constructed meshes  $\{\mathbb{T}_i, \dots, \mathbb{T}_0\}$ . Then, update  $v_i^{(j+1)} = v_i^{(j)} + \lambda_i \delta v_i^{(j)}$  ( $\lambda_i \in (0, 1]$  a relaxation parameter), set  $j = j + 1$  and go back to (2). This process is repeated until a limit  $v_i \in V_i$ , is reached with a certain required accuracy.

(3) *Error estimation:* Solve the (linearized) discrete dual problem

$$z_i \in V_i: \quad A'(v_i; \phi, z_i) = J(\phi) \quad \forall \phi \in V_i,$$

and evaluate the a posteriori error estimate

$$|J(e_i)| \approx \tilde{\eta}(v_i).$$

For controlling the reliability of this bound, i.e. the accuracy in the determination of the dual solution  $z$ , one may check whether  $\|z_i - z_{i-1}\|$  is sufficiently small; if this is not the case, additional global mesh refinement is advisable. If  $\tilde{\eta}(v_i) \leq TOL$  or  $N_i \geq N_{max}$ , then stop. Otherwise cell-wise mesh adaptation yields the new mesh  $\mathbb{T}_{i+1}$ . Then, set  $i = i + 1$  and go back to (1).

This nested solution process is employed in the application presented below. Notice that the derivation of the a posteriori error estimate (3) involves only the solution of *linearized* problems. Hence, the whole error estimation may amount only to a relatively small fraction of the total cost for the solution process.

#### 7.4 Application to the Navier-Stokes equations

The results in this section are collected from Becker [7, 9]; see also [12]. We consider the stationary Navier-Stokes equations

$$-\nu\Delta v + v \cdot \nabla v + \nabla p = 0, \quad \nabla \cdot v = 0, \quad (39)$$

in a bounded domain  $\Omega \subset \mathbb{R}^2$ , with boundary conditions as described in Section 2,

$$v|_{\Gamma_{rigid}} = 0, \quad v|_{\Gamma_{in}} = v^{in}, \quad \nu\partial_n v - pn|_{\Gamma_{out}} = 0.$$

As an example, we consider the flow around the cross section of a cylinder in a channel shown in Figure 33. This is part of a set of benchmark problems discussed in Schäfer/Turek [81].

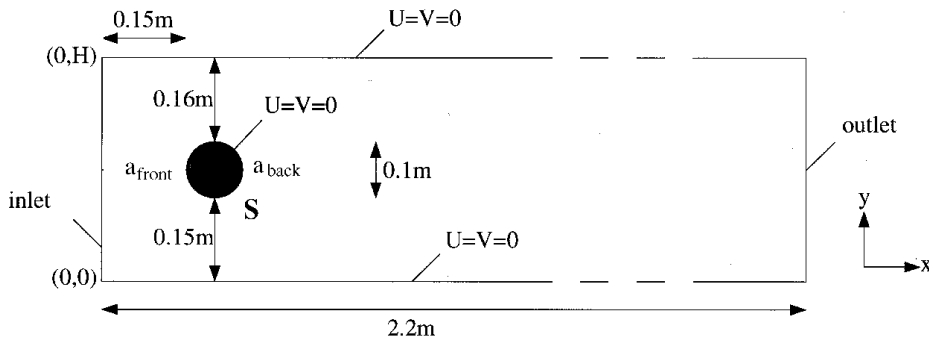


Figure 33: Configuration of the benchmark “flow around a cylinder”

Quantities of physical interest are, for example,

$$\begin{aligned}
\text{pressure drop:} \quad & J_{\Delta p}(v, p) = p(a_{front}) - p(a_{back}), \\
\text{drag coefficient:} \quad & J_{drag}(v, p) = \frac{2}{\bar{U}^2 D} \int_S n \cdot \sigma(v, p) e_x ds, \\
\text{lift coefficient:} \quad & J_{lift}(v, p) = \frac{2}{\bar{U}^2 D} \int_S n \cdot \sigma(v, p) e_y ds,
\end{aligned}$$

where  $S$  is the surface of the cylinder,  $D$  its diameter,  $e_x$  and  $e_y$  the cartesian unit vectors,  $\bar{U}$  the reference velocity, and  $\sigma(v, p) = \frac{1}{2}\nu(\nabla v + \nabla v^T) + pI$  the stress force acting on  $S$ . In our example, the Reynolds number is  $Re = \bar{U}^2 D / \nu = 20$ , such that the flow is stationary. For evaluating the drag and lift coefficients, one may use another representation obtained by the Stokes formula, e.g., for the drag:

$$J_{drag} := \frac{2}{\bar{U}^2 D} \int_S n \cdot \sigma(v, p) e_x ds = \frac{2}{\bar{U}^2 D} \int_{\Omega} \{\sigma(v, p) \nabla \bar{e}_x + \nabla \sigma(v, p) \cdot \bar{e}_x\} dx,$$

where  $\bar{e}_x$  is an extension of  $e_x$  to the interior of  $\Omega$  with support along  $S$ ; see Giles, et al. [28] and Becker [9]. This representation in terms of a domain integral is more robust and accurate than the original one involving a contour integral.

The discretization is by the finite element Galerkin method using the conforming  $Q_1/Q_1$  Stokes element described in Section 3 with least-squares pressure stabilization and streamline diffusion stabilization for the transport. In order to incorporate this scheme in the abstract framework described above, we rewrite it in a more compact form. To this end, we introduce the Hilbert-spaces  $\mathbf{V} := \mathbf{H} \times L$  of pairs  $u := \{v, p\}$  and their discrete analogues  $\mathbf{V}_h := \mathbf{H}_h \times L_h$  of pairs  $u_h := \{v_h, p_h\}$ . Accordingly the Navier-Stokes equations can be written in vector form as follows:

$$Lu := \begin{bmatrix} -\nu \Delta v + v \cdot \nabla v + \nabla p \\ \nabla \cdot v \end{bmatrix} = \begin{bmatrix} f \\ 0 \end{bmatrix}.$$

Further, for  $u := \{v, p\}$  and  $\phi = \{\psi, \chi\}$ , we define the semi-linear form

$$A(u; \phi) := \nu(\nabla v, \nabla \psi) + (v \cdot \nabla v, \psi) - (p, \nabla \cdot \psi) + (\nabla \cdot v, \chi),$$

and the linear functional  $F(\phi) := (f, \psi)$ . Then, the stationary version of the variational formulation (8) is written in the following compact form: Find  $u \in \mathbf{V} + (v_h^{in}, 0)^T$ , such that

$$A(u; \phi) = F(\phi) \quad \forall \phi \in \mathbf{V}. \quad (40)$$

Using the weighted  $L^2$ -bilinear form

$$(v, w)_h := \sum_{K \in \mathbb{T}_h} \delta_K (\nabla v, \nabla w)_K,$$

the stabilized finite element approximation reads as follows: Find  $u_h \in \mathbf{V}_h + (v^{in}, 0)^T$  such that

$$A(u_h; \phi_h) + (Lu_h, S\phi_h)_h = (F, \phi_h) + (F, S\phi_h)_h \quad \forall \phi_h \in \mathbf{V}_h, \quad (41)$$

where the stabilization operator  $S$  is defined by

$$S\phi := \begin{bmatrix} \nu \Delta \psi + v \cdot \nabla \psi + \nabla \chi \\ 0 \end{bmatrix},$$

with the parameter  $\delta$  specified by (36). This formulation contains the pressure and transport stabilization as described in Section 3.

The question is now how to construct a mesh as economical as possible on which the quantities  $J_{\Delta p}(v, p)$ ,  $J_{drag}(v, p)$  and  $J_{lift}(v, p)$  can be computed to the required accuracy, say, of 1%. The a priori design of such a mesh is a difficult task as will be demonstrated by the results of numerical tests below; Table 34 shows a collection of possible *a priori* meshes.

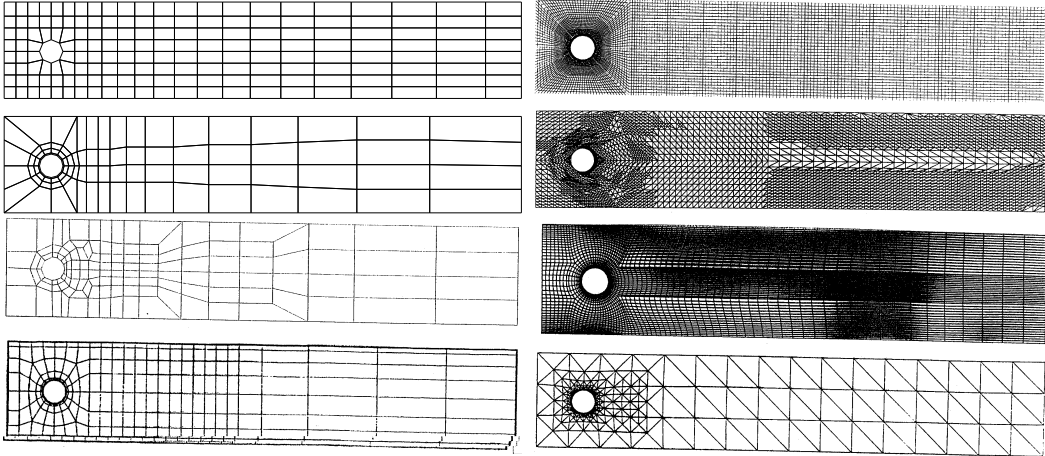


Figure 34: Examples of meshes designed for the benchmark problem “flow around a cylinder”; the first three meshes on the left, “Grid 1”, “Grid 2”, and “Grid 3”, are coarse initial meshes which are to be uniformly refined.

Now, we will discuss the use of *a posteriori* techniques for constructing economical meshes. We denote the discretization error for the pressure by  $e_p := p - p_h$  and that for the velocity by  $e_v := v - v_h$ . By standard arguments relying on the coerciveness properties of the Fréchet derivative of the operator  $L$ , one derives the following energy-norm *a posteriori* error estimate

$$\begin{aligned} \|\nabla e_v\| + \|e_p\| \leq c_{ICS} \left( \sum_{K \in \mathcal{T}_h} \left\{ (h_K^2 + \delta_K) \|R(u_h)\|_K^2 + \right. \right. \\ \left. \left. + \|\nabla \cdot v_h\|_K^2 + \nu h_K \|n \cdot [\nabla v_h]\|_{\partial K}^2 \right\} + \dots \right)^{1/2}, \end{aligned} \quad (42)$$

with the residual  $R(u_h) := \nu \Delta v_h - v_h \cdot \nabla v_h - \nabla p_h$ . In this estimate the “...” stand for additional terms representing the errors in approximating the inflow data and the curved boundary  $S$ ; they can be expected to be small compared to the other residual terms and are usually neglected. In this estimate the *interpolation constant*  $c_I$  can be determined and is of moderate size  $c_I \sim 0.2$ . The most critical point is the *stability constant*  $c_S$  which is completely unknown. It is related to the constant in the coerciveness estimate of the tangent form  $A'(v; \cdot, \cdot)$  of  $A(\cdot; \cdot)$  taken at the solution  $v$ ,

$$\|\nabla w\| + \|q\| \leq c_S \sup_{\phi \in V} \left\{ \frac{A'(v; z, \phi)}{\|\nabla \psi\| + \|\chi\|} \right\},$$

where  $z = \{w, q\}$  and  $\phi = \{\psi, \chi\}$ . In order to use this error bound for mesh-size control, we have set it to  $c_S = 1$ .

The error estimate (42) is not appropriate for controlling the error in local quantities like drag and lift since it measures the residual uniformly over the whole computational domain. One way of introducing more a priori information into the mesh refinement process based on (42) is to start from an initial mesh which is already refined towards the contour  $S$ . Alternatively, one may also use (on heuristic grounds) additional weighting factors which enforce stronger mesh refinement in the neighborhood of  $S$ . The resulting global error indicator reads as follows:

$$\begin{aligned} \|\nabla e_v\| + \|e_p\| \leq c_I c_S \left( \sum_{K \in \mathbb{T}_h} \sigma_K \left\{ (h_K^2 + \delta_K) \|R(u_h)\|_K^2 \right. \right. & (43) \\ \left. \left. + \|\nabla \cdot v_h\|_K^2 + \nu h_K \|n \cdot [\nabla v_h]\|_{\partial K}^2 \right\} + \dots \right)^{1/2}, \end{aligned}$$

where the weights  $\sigma_K$  are chosen large along  $S$ .

Correctly *weighted a posteriori* error estimates can be obtained following the general line of argument described above. The approximate dual problem seeks  $z := \{w, q\} \in V$  satisfying

$$A'(u_h; \varphi, z) + (L'(u_h)^* \varphi, Sz)_\delta = J(\varphi) \quad \forall \varphi \in V, \quad (44)$$

where  $A'(u_h; \cdot, \cdot)$  and  $L'(u_h)^*$  are the tangent form and adjoint tangent operator of  $A(\cdot; \cdot)$  and  $L(\cdot)$ , respectively. The resulting weighted a posteriori estimate for the error  $e := u - u_h$  becomes

$$|J(e)| \leq \sum_{K \in \mathbb{T}_h} \left\{ \rho_K \omega_K + \rho_{\partial K} \omega_{\partial K} + \rho_K^{div} \omega_K^{div} + \dots \right\}, \quad (45)$$

with the local residual terms and weights defined by

$$\begin{aligned} \rho_K &= \|R(u_h)\|_K, & \omega_K &= \|w - w_h\|_K + \delta_K \|v_h \cdot \nabla(w - w_h) + \nabla(q - q_h)\|_K, \\ \rho_{\partial K} &= \frac{1}{2} \nu \|n \cdot [\nabla v_h]\|_{\partial K}, & \omega_{\partial K} &= \|w - w_h\|_{\partial K}, \\ \rho_K^{div} &= \|\nabla \cdot v_h\|_K, & \omega_K^{div} &= \|q - q_h\|_K. \end{aligned}$$



The dots “...” stand again for additional terms measuring the errors in approximating the inflow and the curved cylinder boundary. For more details on this aspect, we refer to [14] and Becker [9]. The bounds for the dual solution  $z = \{w, q\}$  are obtained computationally by replacing the unknown solution  $u$  in the convection term by its approximation  $u_h$  and solving the resulting linearized problem on the same mesh. From this approximate dual solution  $\tilde{z}_h$ , patchwise biquadratic interpolations are taken to approximate  $z$  in evaluating the weights  $\omega_K^{(i)}$ ,  $I_h^{(2)}\tilde{z}_h - z_h \approx z - z_h$ . This avoids the occurrence of interpolation constants.

Table 4 shows the corresponding results for the pressure drop computed on four different types of meshes:

- (i) Hierarchically refined meshes starting from coarse meshes of type “Grid 1” and “Grid 2” as shown in Figure 34.
- (ii) Adapted meshes using the global energy-norm error estimate (42) with enforced refinement along the contour  $S$ ; see Figure 43.
- (iii) Adapted meshes using the weighted error estimate (45) for the pressure drop; see Figure 42.

These results demonstrate clearly the superiority of the weighted error estimate (45) in computing local quantities. It produces an error of less than 1% already after 6 refinement cycles on a mesh with less than 1400 unknowns while the other algorithms use more than 21000 unknowns to achieve the same accuracy (the corresponding values are printed in boldface). Corresponding sequences of meshes generated by the weighted energy error estimate (42) and the energy-error estimate (43) are seen in Figures 35 and 36.

Table 5 contains some results of the computation of drag and lift coefficients using the corresponding weighted error estimates. The effectivity index is defined by  $I_{eff} := \eta(u_h)/|J(e)|$ . Finally, Figure 37 shows plots of the dual solutions occurring in the computation of pressure drop, drag and lift.

Table 4: Results of the pressure drop computation (ref. value  $\Delta p = 0.11752016\dots$ ); a) upper row: on uniformly refined meshes of type Grid1 and Grid2, b) lower row: on adaptively refined meshes starting from a coarse mesh Grid1; from Becker [7].

Uniform Refinement, Grid1			Uniform Refinement, Grid2		
$L$	$N$	$\Delta p$	$L$	$N$	$\Delta p$
1	2268	0.109389	1	1296	0.106318
2	8664	0.110513	2	4896	0.112428
3	33840	0.113617	3	19008	0.115484
4	133728	0.115488	4	<b>74880</b>	<b>0.116651</b>
5	<b>531648</b>	<b>0.116486</b>	5	297216	0.117098

Adaptive Refinement, Grid1			Weighted Adaptive Refinement		
$L$	$N$	$\Delta p$	$L$	$N$	$\Delta p$
2	1362	0.105990	4	650	0.115967
4	5334	0.113978	6	<b>1358</b>	<b>0.116732</b>
6	<b>21546</b>	<b>0.116915</b>	9	2858	0.117441
8	86259	0.117379	11	5510	0.117514
10	330930	0.117530	12	8810	0.117527

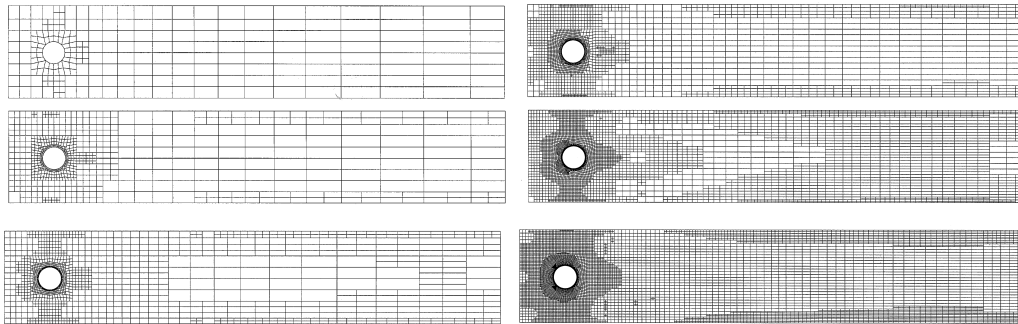


Figure 35: A sequence of refined meshes generated by the (heuristically) weighted global energy estimate; from Becker [7].

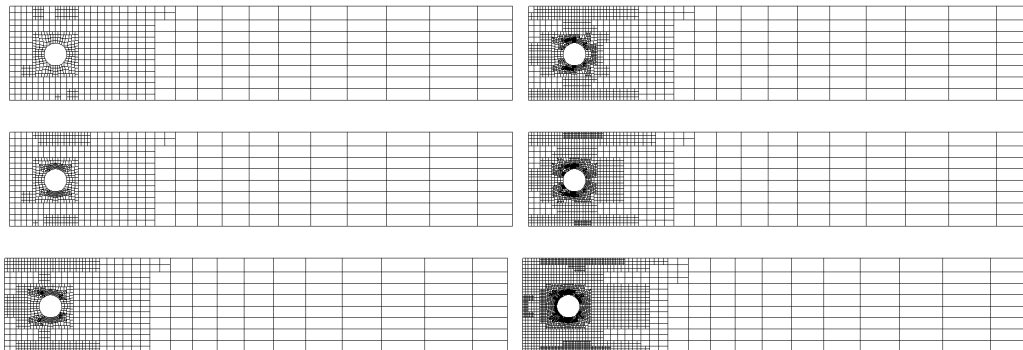


Figure 36: A sequence of refined meshes generated by the weighted error estimate for the pressure drop; from Becker [7].

Table 5: Results of the cylinder flow computations of drag and lift (ref. values  $c_{drag} = 5.579535\dots$  and  $c_{lift} = 0.0106189\dots$ ) on adaptively refined meshes starting from a coarse mesh of type Grid1; from Becker [9].

Computation of drag					Computation of lift				
$L$	$N$	$c_{drag}$	$\eta_{drag}$	$I_{eff}$	$L$	$N$	$c_{lift}$	$\eta_{lift}$	$I_{eff}$
3	251	5.780186	$2.0e-1$	0.5	3	296	0.007680	$2.9e-3$	5.0
4	587	5.637737	$5.8e-2$	0.6	4	764	0.009249	$1.4e-3$	5.0
5	1331	5.568844	$1.0e-2$	1.6	5	1622	0.009916	$7.3e-4$	5.0
6	3953	5.576580	$2.5e-3$	2.0	6	4466	0.010144	$5.0e-4$	2.5
7	8852	5.578224	$8.7e-4$	2.5	7	8624	0.010267	$3.8e-5$	2.0
8	16880	5.578451	$6.5e-4$	1.6	8	18093	0.010457	$1.9e-5$	2.0
9	34472	5.578883	$2.1e-4$	2.0	9	34010	0.010524	$1.2e-4$	1.6

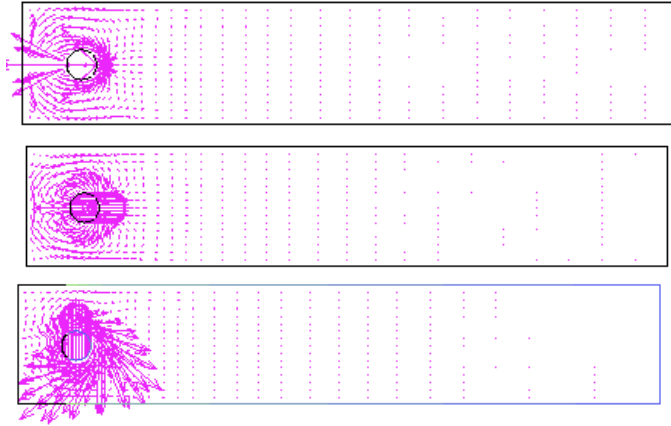


Figure 37: Velocity plots of the dual solution for pressure drop (top), drag (middle), and lift (bottom); from Becker [9].

## 7.5 The nonstationary case

The extension of the approach to mesh adaptivity described above to the *nonstationary* Navier-Stokes equations is presently under development.

(I) The traditional method for a posteriori time-step selection is based on the concept of controlling the *local* “truncation error” but neglecting the *global* error accumulation. In its simplest form this strategy uses the condition

$$\frac{1}{3}c_S k_n^2 (U_{k/2}^n - U_k^n) \approx TOL, \quad (46)$$

where  $U_k^n$  and  $U_{k/2}^n$  are the solutions computed from the preceding approximation  $U^{n-1}$  at  $t_{n-1}$  by a second-order scheme (e.g. the Crank-Nicolson scheme) with time-step sizes  $k$  and  $k/2$ , respectively. For a more detailed description of techniques of this type, we refer to Turek [97].

(II) The extension of the residual-based error control described above to non-stationary problems is based on a time discretization which has also the features of a Galerkin method. These are for example the so-called “continuous” or “discontinuous” Galerkin methods of polynomial degree  $r \geq 0$  (“cG(r)” or “dG(r)” methods). The lowest-order examples are the dG(0) method which (in the autonomous case) is equivalent to the backward Euler scheme, the dG(1) method which is similar to an implicit Runge-Kutta scheme of third order, and the cG(1) method which can be interpreted as a variant of the Crank-Nicolson scheme. In particular, the dG(1) method is attractive for solving the nonstationary Navier-Stokes problem because of its superior accuracy (compared to the dG(0) method).

The result of error estimation using a duality argument is an a posteriori error estimate of the form (see [55] and Hartmann [39])

$$\|U_k^n - u(\cdot, t_n)\| \approx \sum_{m=1}^n \{k_m^3 \omega_m \|d_t U_k^m\| + \dots\} \quad (47)$$

where  $d_t U_k^m = k_m^{-1}(U_k^m - U_k^{m-1})$  are the time-difference quotients of the computed solution and  $\omega_m$  are weighting factors obtained by solving a “backward in time” space-time dual problem. The dots “...” refer to residual terms of the spatial discretization. The main problem with this approach is its huge computational work; in a nonlinear problem the “forward” solution  $\{U_k^m\}_{m=0}^n$  enters the linearized dual problem as coefficient and needs therefore to be stored over the whole time interval. Moreover, in this way error control can be achieved only at single times  $t_n$  or for the time-averaged error. Controlling the error uniformly in time requires (theoretically) to solve a dual problem at each discrete time level resulting in prohibitively high cost. The economical realization of this concept for computing nonstationary flows involving global error control is still an open problem.

**Open Problem 7.3:** *Devise a strategy for adapting the stabilization parameters  $\delta_K$  simultaneously with the mesh size  $h_K$  on the basis of the a posteriori error estimate (45).*

**Open Problem 7.4:** *Derive an a posteriori error estimate of the form (45) for the full space-time discretization of the Navier-Stokes equations and devise a strategy for simultaneous adaptation of mesh sizes  $h_K$  and time steps  $k_n$ .*

## 8 Extension to weakly compressible flows

In this last section, we discuss the extension of the computational methodology described above to certain *compressible* flows. The flows of interest are those in which density changes are induced by temperature gradients resulting for example from heat release by chemical reactions. Such “weakly” compressible flows are characterized by low-Mach-number speed and hydrodynamically incompressible behavior. Here, the dominant problem is that of stiff velocity-pressure coupling while shocks or large pressure gradients do not develop. We recall the system of conservation equations for mass, momentum and energy, in the case of a stationary flow:

$$\nabla \cdot [\rho v] = 0, \quad (1)$$

$$\rho v \cdot \nabla v - \nabla \cdot [\mu \nabla v + \frac{1}{3} \mu \nabla \cdot v I] + \nabla p_{tot} = \rho f, \quad (2)$$

$$c_p \rho v \cdot \nabla T - \nabla \cdot [\lambda \nabla T] = h. \quad (3)$$

Here, again  $v$  is the velocity,  $\rho$  the density,  $p_{tot}$  the (total) pressure and  $T$  the temperature of the fluid occupying a two- or three-dimensional region  $\Omega$ . The dynamic viscosity  $\mu > 0$ , the heat capacity  $c_p > 0$ , the heat conductivity  $\lambda$ , the external volume force  $f$  and the heat source  $h$  are given. Since we only consider low-speed flows, the influence of stress and hydrodynamic pressure in the energy equation can be neglected. In general,  $f$  as well as  $h$  implicitly depend on the temperature  $T$  and on further quantities describing the release of heat for example through chemical reactions. Here, we will simply consider the heat source  $h$  as given. The coupling between pressure and density is assumed as that of a perfect gas,

$$p_{tot} = R \rho T, \quad (4)$$

where  $R$  is the gas constant. As mentioned above, we consider hydrodynamically incompressible flows. Accordingly, the pressure is split into two parts,

$$p_{tot}(x, t) = \bar{p}(t) + p_{hyd}(x, t),$$

namely the spatial mean value

$$\bar{p} := |\Omega|^{-1} \int_{\Omega} p_{tot}(x, t) dx,$$

and the “hydrodynamic pressure”  $p_{hyd}(x, t)$ . In a weakly compressible flow, the pressure variation due to hydrodynamic mechanisms is assumed to be small compared to the mean value of the total pressure,

$$|p_{hyd}| \ll |\bar{p}|,$$

which is determined by thermodynamic effects. Accordingly, we call  $P_{th}(t) = \bar{p}(t)$  the “thermodynamic pressure”. In the “low-Mach-number approximation” the hydrodynamic pressure occurs in the momentum equation

$$\rho v \cdot \nabla v - \nabla \cdot [\mu \nabla v + \frac{1}{3} \mu \nabla \cdot v I] + \nabla p_{hyd} = \rho f, \quad (5)$$

while the pressure-density coupling in the equation of state (4) is expressed in terms of the “thermodynamic pressure”

$$\rho = \frac{P_{th}}{RT}. \quad (6)$$

In many applications, this set of equations has to be supplemented by further conservation equations for species concentrations and complicated nonlinear source terms representing the chemical reactions. Here, we restrict ourselves to the simple case of low-Mach-number flow, where temperature variations are induced by outer source terms. The thermodynamic pressure  $P_{th}(t)$  is supposed to be determined by a priori considerations; for more details, see Braack [17] and also [18].

**Open Problem 8.1:** *Estimate the error caused by neglecting stress and hydrodynamic pressure in the energy equation. Prove corresponding error bounds for the low-Mach-number approximation in terms of the Mach number.*

Since in the above approximation the density occurs as a secondary variable determined by the temperature through the equation of state, it appears natural to use the pressure  $p := p_{hyd}$  together with the velocity  $v$  and the temperature  $T$  as primal variables in the computational model. We use the equation of state to rewrite the continuity equation as an equation for velocity and temperature:

$$\nabla \cdot v - T^{-1} v \cdot \nabla T = 0. \quad (7)$$

Furthermore, introducing the modified pressure  $p := p_{hyd} - \frac{1}{3}\mu \nabla \cdot v$ , the momentum equation can be written as

$$\rho v \cdot \nabla v - \nabla \cdot [\mu \nabla v] + \nabla p = \rho f, \quad (8)$$

while the energy equation keeps the form

$$c_p \rho v \cdot \nabla T - \nabla \cdot [\lambda \nabla T] = h. \quad (9)$$

The temperature-dependent functions  $\mu = \mu(T)$  and  $c_p = c_p(T)$  are usually given in terms of polynomial fits from data bases. The density  $\rho$  is expressed by the algebraic relation (6) in terms of the temperature. The system is closed by imposing appropriate boundary conditions,

$$v|_{\Gamma_{rigid}} = 0, \quad v|_{\Gamma_{in}} = v^{in}, \quad \mu \partial_n v + p n|_{\Gamma_{out}} = 0, \quad T|_{\partial\Omega} = \hat{T}, \quad (10)$$

where again  $\Gamma_{rigid}$ ,  $\Gamma_{in}$ ,  $\Gamma_{out}$  are the rigid part, the inflow part and the outflow part of the boundary  $\partial\Omega$ , respectively. For questions of well-posedness of this type of problem, we refer to the relevant literature, e.g., Feistauer [27] and Lions [60].

The starting point for a finite element discretization of problem (7), (8), (9), and (6) is again its variational formulation. To formulate this, we introduce the natural function spaces as already used above,

$$L \subset L^2(\Omega), \quad \mathbf{H} \subset H^1(\Omega)^d, \quad R \subset H^1(\Omega).$$

for the pressure  $p \in L$ , the velocity  $v \in \mathbf{H}$ , and the temperature  $T \in R$ . For a compact notation, we set  $V := L \times \mathbf{H} \times R$ . Prescribed Dirichlet data  $\hat{v}$  and  $\hat{T}$  can be included by seeking the weak solutions in appropriate sub-manifolds,

$$p \in L, \quad v \in \hat{v} + \mathbf{H}, \quad T \in \hat{T} + R.$$

Then, the triple  $u := \{p, v, T\}$  is determined by the variational equations

$$(\nabla \cdot v, \chi) - (T^{-1}v \cdot \nabla T, \chi) = 0, \quad \forall \chi \in L, \quad (11)$$

$$(\rho v \cdot \nabla v, \psi) + (\mu \nabla v, \nabla \psi) - (p, \nabla \cdot \psi) = (\rho f, \psi) \quad \forall \psi \in \mathbf{H}, \quad (12)$$

$$(\rho c_p v \cdot \nabla T, \pi) + (\lambda \nabla T, \nabla \pi) = (h, \pi) \quad \forall \pi \in R. \quad (13)$$

In the following analysis, we consider for simplicity only the case of pure Dirichlet boundary conditions. In this case the pressure is determined only modulo constants and the corresponding solution space is  $L = L_0^2(\Omega)$ .

Now, the finite element discretization replaces the (infinite dimensional) function spaces  $L$ ,  $\mathbf{H}$ , and  $R$  by finite dimensional *discrete* spaces denoted by  $L_h$ ,  $\mathbf{H}_h$ , and  $R_h$ . Here, we think of finite element spaces based for example on conforming  $Q_1$  approximation for all physical quantities. The corresponding *discrete* solutions  $p_h \in L_h$ ,  $v_h \in \hat{v}_h + \mathbf{H}_h$ , and  $T_h \in \hat{T}_h + R_h$  are determined through the system

$$(\nabla \cdot v_h, \chi_h) - (T_h^{-1}v_h \cdot \nabla T_h, \chi_h) = 0, \quad \forall \chi_h \in L_h, \quad (14)$$

$$(\rho v_h \cdot \nabla v_h, \psi_h) + (\mu \nabla v_h, \nabla \psi_h) - (p_h, \nabla \cdot \psi_h) = (\rho f, \psi_h) \quad \forall \psi_h \in \mathbf{H}_h, \quad (15)$$

$$(\rho c_p v_h \cdot \nabla T_h, \pi_h) + (\lambda \nabla T_h, \nabla \pi_h) = (h, \pi_h) \quad \forall \pi_h \in R_h, \quad (16)$$

with coefficients  $\mu = \mu(T_h)$  and  $c_p = c_p(T_h)$ . The compact formulation of the system (14)-(16) makes use of the semi-linear form

$$\begin{aligned} A(u; \phi) := & (\nabla \cdot v, \chi) - (T^{-1}v \cdot \nabla T, \chi) + (\rho v \cdot \nabla v, \psi) + (\mu \nabla v, \nabla \psi) \\ & - (p, \nabla \cdot \psi) + (\rho c_p v \cdot \nabla T, \pi) + (\lambda \nabla T, \nabla \pi), \end{aligned}$$

and the linear form

$$F(\phi) = (\rho f, \psi) + (h, \pi),$$

defined for triples  $u = \{p, v, T\}$ ,  $\phi = \{\chi, \psi, \pi\} \in V$ . With this notation, the problem reads as follows: Find  $u \in \hat{u} + V$ , such that

$$A(u; \phi) = F(\phi) \quad \forall \phi \in V. \quad (17)$$

where  $\hat{u}$  represents Dirichlet boundary data for all components. The corresponding discrete problem reads: Find  $u_h \in \hat{u}_h + V_h$ , such that

$$A(u_h; \phi_h) = F(\phi_h) \quad \forall \phi_h \in V_h. \quad (18)$$

In general this system is unstable and needs stabilization with respect to the stiff velocity-pressure coupling as well as the transport terms.

## 8.1 Least-squares stabilization

The stabilization is introduced into the system (18) by using pressure stabilization and streamline diffusion as discussed above in the context of the incompressible Navier-Stokes equations. The corresponding stabilization terms are listed below:

- Pressure stabilization:

$$\begin{aligned} s_h^p(u_h, \chi_h) &= \sum_{K \in \mathcal{T}_h} \alpha_K (\bar{v}_h \cdot \nabla v_h - \nabla \cdot [\mu \nabla v_h] + \nabla p_h, \nabla \chi_h)_K, \\ r_h^p(u_h, \chi_h) &= \sum_{K \in \mathcal{T}_h} \alpha_K (\rho f, \nabla \chi_h)_K. \end{aligned}$$

- Streamline diffusion for the velocities:

$$\begin{aligned} s_h^v(u_h, \psi_h) &= \sum_{K \in \mathcal{T}_h} \delta_K (\rho \bar{v}_h \cdot \nabla v_h - \nabla \cdot [\mu \nabla v_h] + \nabla p_h, \rho \bar{v}_h \cdot \nabla \psi_h)_K, \\ r_h^v(u_h, \psi_h) &= \sum_{K \in \mathcal{T}_h} \delta_K (\rho f, \rho \bar{v}_h \cdot \nabla \psi_h)_K. \end{aligned}$$

- Streamline diffusion for the temperature:

$$\begin{aligned} s_h^T(u_h, \pi_h, \chi_h) &= \sum_{K \in \mathcal{T}_h} \gamma_K (\rho c_p \bar{v}_h \cdot \nabla T_h - \nabla \cdot [\lambda \nabla T_h], \rho c_p \bar{v}_h \cdot \nabla \pi_h)_K, \\ r_h^T(u_h, \pi_h, \chi_h) &= \sum_{K \in \mathcal{T}_h} \gamma_K (h, \rho c_p \bar{v}_h \cdot \nabla \pi_h)_K. \end{aligned}$$

Here,  $\bar{v}_h$  is a suitable approximation to the current velocity field  $v_h$ , taken for example from a preceding iteration step. We denote the sum over these  $h$ -dependent stabilization terms by  $s_h(\cdot, \cdot)$  and  $r_h(\cdot)$ , respectively,

$$\begin{aligned} s_h(u_h, \phi) &:= s_h^p(u_h, \chi) + s_h^v(u_h, \psi) + s_h^T(u_h, \pi), \\ r_h(u_h, \phi) &:= r_h^p(u_h, \chi) + r_h^v(u_h, \psi) + r_h^T(u_h, \pi). \end{aligned}$$

Then, with  $A_h(\cdot; \cdot) := A(\cdot; \cdot) + s_h(\cdot, \cdot)$  and  $F_h(\cdot) := F(\cdot) + r_h(\cdot)$ , the discrete equations can be written in compact form

$$A_h(u_h; \phi) = F_h(\phi_h) \quad \forall \phi \in V_h. \quad (19)$$

In order to ensure symmetry for the resulting stabilized system,  $\alpha_K$  should be taken equal to  $\delta_K$ . The stability and consistence of this formulation can be analyzed by similar techniques as used in the case of the incompressible Navier-Stokes equations; see Braack [17]. One obtains the following condition for the parameters  $\delta_K$ :

$$\delta_K = \left[ \frac{\mu}{h_K^2} + \frac{|\rho \bar{v}_h|_\infty}{h_K} \right]^{-1}, \quad \gamma_K = \left[ \frac{\lambda}{h_K^2} + \frac{|\rho c_p \bar{v}_h|_\infty}{h_K} \right]^{-1}. \quad (20)$$



**Open Problem 8.2:** Derive a formula for the stabilization parameters  $\delta_K$  and  $\gamma_K$  which leads to a robust scheme on general meshes with large aspect ratio  $\sigma_h$ .

## 8.2 Computational approach

For solving the model for weakly compressible flow introduced above, we want to use the methodology discussed for incompressible flows.

(i) *Explicit defect correction coupling:*

The simplest way to use an “incompressible solver” for computing weakly compressible flows is by a defect correction iteration. The step  $\{p_h^{l-1}, v_h^{l-1}, T_h^{l-1}\} \rightarrow \{p_h^l, v_h^l, T_h^l\}$  of this scheme proceeds as follows:

1. The nonlinear coefficients (density, transport vectors, etc.) are frozen at  $\{p_h^{l-1}, v_h^{l-1}, T_h^{l-1}\}$ . Corresponding corrections  $\{\delta p_h^l, \delta v_h^l, \delta T_h^l\} \in V_h$  are determined by solving the linearized system:

$$\begin{aligned} (\nabla \cdot \delta v_h^l, \chi_h) &= (d_p^{l-1}, \chi_h), \quad \forall \chi_h \in L_h, \\ (\rho v_h^{l-1} \cdot \nabla \delta v_h^l, \phi_h) + (\mu \nabla \delta v_h^l, \nabla \phi_h) - (\delta p_h^l, \nabla \cdot \phi_h) &= (d_v^{l-1}, \phi_h) \quad \forall \phi_h \in \mathbf{H}_h, \\ (\rho c_p v_h^{l-1} \cdot \nabla \delta T_h^l, \pi_h) + (\lambda \nabla \delta T_h^l, \nabla \pi_h) &= (d_t^{l-1}, \pi_h) \quad \forall \pi_h \in R_h, \end{aligned}$$

where  $d_p^{l-1}$ ,  $d_v^{l-1}$ , and  $d_t^{l-1}$  are the defects of the iteration  $\{p_h^{l-1}, v_h^{l-1}, T_h^{l-1}\}$ . For the sake of robustness, the pressure and transport stabilization described above has also to be applied to this problem.

2. The new solution vector is obtained by

$$p_h^l = p_h^{i-1} + \kappa_l \delta p_h^l, \quad v_h^l = v_h^{i-1} + \kappa_l \delta v_h^l, \quad T_h^l = T_h^{i-1} + \kappa_l \delta T_h^l,$$

with some relaxation parameter  $\kappa_l \in (0, 1]$ , and the density is updated according to  $\rho_h^l = P_{th} / (RT_h^l)$ .

3. The iteration is continued until some stopping criterion is satisfied.

In each step of this iteration a linearized Navier-Stokes problem supplemented by a heat transfer equation is to be solved. This may be accomplished by using the methods described above for the incompressible Navier-Stokes equations. Hence, the “incompressible solver” is used for preconditioning the defect correction iteration for solving the full system (18). However, this simple defect correction process may converge very slowly in the case of large temperature gradients (e.g., caused by strong heat release in chemical reactions). This lack of robustness can be cured by making the iteration more implicit.

(ii) *Semi-implicit defect correction coupling:*

In order to achieve better control on the variation of temperature, one may use the following more implicit iteration:

1. The nonlinear coefficients (density, transport vectors, etc.) are frozen at  $\{p_h^{l-1}, v_h^{l-1}, T_h^{l-1}\}$ . Corresponding corrections  $\{\delta p_h^l, \delta v_h^l, \delta T_h^l\} \in V_h$  are then determined by solving the linearized system:

$$\begin{aligned} (\nabla \cdot \delta v_h^l, \chi_h) + ((T_h^{l-1})^{-1} v_h^{l-1} \cdot \nabla \delta T_h^l, \chi) &= (d_p^{l-1}, \chi_h) \quad \forall \chi_h \in L_h, \\ (\rho v_h^{l-1} \cdot \nabla \delta v_h^l, \phi_h) + (\mu \nabla \delta v_h^l, \nabla \phi_h) - (\delta p_h^l, \nabla \cdot \phi_h) &= (d_v^{l-1}, \phi_h) \quad \forall \phi_h \in \mathbf{H}_h, \\ (\rho c_p v_h^{l-1} \cdot \nabla \delta T_h^l, \pi_h) + (\lambda \nabla \delta T_h^l, \nabla \pi_h) &= (d_t^{l-1}, \pi_h) \quad \forall \pi_h \in R_h, \end{aligned}$$

where  $d_p^{l-1}$ ,  $d_v^{l-1}$ , and  $d_t^{l-1}$  are the defects of the iteration  $\{p_h^{l-1}, v_h^{l-1}, T_h^{l-1}\}$ . Again, the pressure and transport stabilization described above has to be applied.

2. The new solution vector is obtained by

$$p_h^l = p_h^{i-1} + \kappa_l \delta p_h^l, \quad v_h^l = v_h^{i-1} + \kappa_l \delta v_h^l, \quad T_h^l = T_h^{i-1} + \kappa_l \delta T_h^l,$$

with some relaxation parameter  $\kappa_l \in (0, 1]$ , and the density is updated according to  $\rho_h^l = P_{th}/(RT_h^l)$ .

3. The iteration is continued until some stopping criterion is satisfied.

This solution method has been used in Braack [17] for the simulation of low-Mach-number combustion processes; see also [10] and [18].

### 8.3 The algebraic system

In each substep of the defect correction iterations described above, we have to solve linear problems for the coefficients  $x_j = \{x_j^{(p)}, x_j^{(v)}, x_j^{(T)}\}$  including the components for pressure, velocity and temperature in the basis representations

$$p_h = \sum_{j=1}^N x_j^{(p)} \psi_j, \quad v_h = \sum_{j=1}^N x_j^{(v)} \psi_j, \quad T_h = \sum_{j=1}^N x_j^{(T)} \psi_j.$$

The system sub-matrices corresponding to the different components are obtained from the coupled system by taking first test functions of the form  $\phi_h = \{\psi_h, 0, 0\}$ :

$$\begin{aligned} A_h(u_h; \{\psi_h, 0, 0\}) &= (\nabla \cdot v_h, \psi_h) - (\bar{T}_h^{-1} \bar{v}_h \cdot \nabla T_h, \psi_h) \\ &\quad + \sum_{K \in \mathcal{T}_h} \delta_K (\rho \bar{v}_h \cdot \nabla v_h - \nabla \cdot (\mu \nabla v_h) + \nabla p_h, \nabla \psi_h)_K. \end{aligned}$$

Analogously, taking the test functions  $\phi_h = \{0, \psi_h, 0\}$ , we obtain the equation for the velocity components,

$$\begin{aligned} A_h(u_h; \{0, \psi_h, 0\}) &= (\rho \bar{v}_h \cdot \nabla v_h, \psi_h) + (\mu \nabla v_h, \nabla \psi_h) - (p_h, \nabla \cdot \psi_h) + \\ &\quad \sum_{K \in \mathcal{T}_h} \delta_K (\rho \bar{v}_h \cdot \nabla v_h - \nabla \cdot (\mu \nabla v_h) + \nabla p_h, \rho \bar{v}_h \cdot \nabla \psi_h)_K, \end{aligned}$$

and by taking the test functions  $\phi_h = (0, 0, \pi_h)$  the equation for the temperature component,

$$A_h(u_h; \{0, 0, \psi_h\}) = (\rho c_p \bar{v}_h \cdot \nabla T_h, \psi_h) + (\lambda \nabla T_h, \nabla \psi_h) + \sum_{K \in \mathcal{T}_h} \gamma_K (\rho c_p \bar{v}_h \cdot \nabla T_h - \nabla \cdot [\lambda \nabla T_h], \rho c_p \bar{v}_h \cdot \nabla \psi_h).$$

Ordering the unknowns in a physically block-wise sense, i.e., marching through the set of nodal points and attaching to each node the corresponding submatrix containing the unknowns of all physical quantities, we obtain “nodal matrices”  $\mathcal{A}_{ij}$  of the form

$$\mathcal{A}_{ij} = \begin{bmatrix} B_{pp} & B_{pv} & B_{pT} \\ B_{vp} & B_{vv} & B_{vT} \\ B_{Tp} & B_{Tv} & B_{TT} \end{bmatrix},$$

where the indices  $p, v, T$  indicate the corresponding contributions. Looking at the equations, we see that almost all physical components are coupled with each other; only the pressure does not appear in the temperature equation, i.e.,  $B_{Tp} = 0$ . Several of the other couplings are of minor importance and may be neglected in building an approximating nodal matrix  $\tilde{\mathcal{A}}_{ij}$  to  $\mathcal{A}_{ij}$ . One could think of a complete decoupling of the flow variables  $\{p, v\}$  from the temperature  $T$  (or other state variables describing for example chemical reactions) resulting in an approximation of the form

$$\tilde{\mathcal{A}}_{ij} = \begin{bmatrix} B_{pp} & B_{pv} & 0 \\ B_{vp} & B_{vv} & 0 \\ 0 & 0 & B_{TT} \end{bmatrix}.$$

However, such a simplification is not appropriate in computing processes in which the temperature has a significant influence on the flow field and vice versa. For example, in combustion problems, density variations are mainly caused by changes of the temperature. A detailed discussion of this issue can be found in Braack [17] and in [18].

## 8.4 An example of chemically reactive flow

We close this section by presenting some results from Braack [17] on computations for low-Mach-number flows with chemical reactions. The configuration considered is the model of a methane burner with a complicated geometry and using a sophisticated reaction mechanism. A stoichiometric mixture of methane  $CH_4$  and air  $O_2/N_2$  flows from the bottom of the burner through a sample of slots of uniform width  $2\text{ mm}$  and three different heights (varying from  $14\text{ mm}$  to  $11\text{ mm}$ ). The columns have a uniform width of  $1.5\text{ mm}$ . The inflow velocity is uniformly  $0.2\text{ m/s}$ . The Reynolds number in this model is about  $Re = 90$ . The geometry is shown in Figure 38.

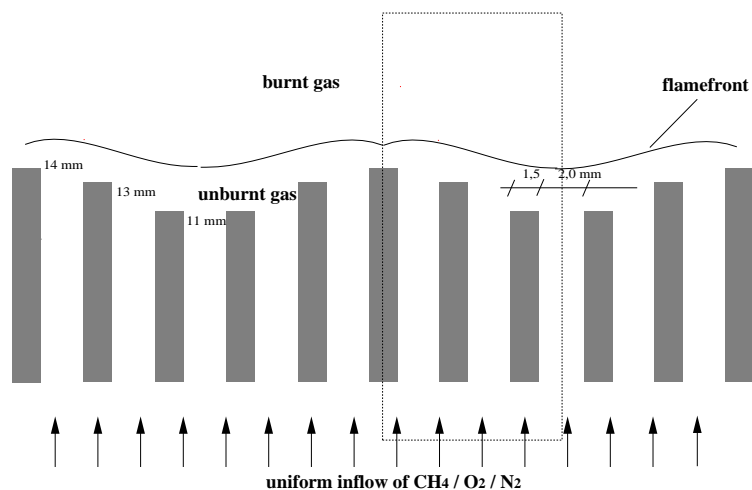


Figure 38: Geometry of a methane burner; from Braack [17].

Due to the heating of the slots, the flow accelerates up to approximately  $1\text{ m/s}$ . Since this is higher than the flame velocity of a stoichiometric methane flame the flame front is located above the slots. For lower inflow velocities, the flame moves downstream into the slots and extinguishes as a result of the heat loss by the cold walls.

If the solution is assumed to be spatially periodic, it is sufficient to restrict the computational domain  $\Omega$  to only three slots, as shown in Figure 39. The boundaries at the left and right hand of  $\Omega$  are symmetry boundary conditions. The walls of the slots are described by Dirichlet conditions for the temperature and Neumann conditions for the species. The calculation on the coarsest mesh (with 1344 cells) uses a time-stepping procedure to provide a physically correct starting value. Then, on the finer meshes the stationary fixed-point defect correction iteration converges. In order to obtain ignition, the temperature for the initial solution is set to  $2000\text{ K}$  at the points above the slots. The reaction mechanism is that of Smooke [80] with 15 species and 84 elementary reactions (42 bidirectional), supplemented by two further species,  $N$  and  $NO$ , and 4 additional reactions to describe their formation.

The solution is obtained on an adaptively refined mesh with refinement criterion based on the linear functional

$$J(u) = |\Omega|^{-1} \int_{\Omega} T dx ,$$

in order to capture the temperature distribution accurately. The finest mesh is shown in Figure 39; we see local mesh refinement at the flame front and below the slots where the velocity field changes. The mesh is automatically adapted and no hand-fitting on the basis of a priori knowledge of the solution is necessary to find the appropriate balance of the mesh-size distribution. The CPU time required for such a simulation with about 5,000 cells ( $\approx 100,000$  unknowns) is approximately 6 hours on a Pentium II (233 Mhz) when the initial guess on the coarse grid with approximately 1300 cells is given.

The computed pressure and the main velocity component are shown in Figure 39. Due to the strong heat release the flow accelerates by a factor of 10 at the outflow of the slots. At the walls of the slots, Dirichlet conditions for the temperature are imposed, varying linearly from 298 K at the bottom up to 393 K, 453 K and 513 K for the three different walls. This leads to a higher outflow velocity at the longer slot compared to the shorter ones. Therefore, the lift-off of the flame is substantially higher at the longer slot, leading to the common Bunsen cone formed by two neighboring longer slots.

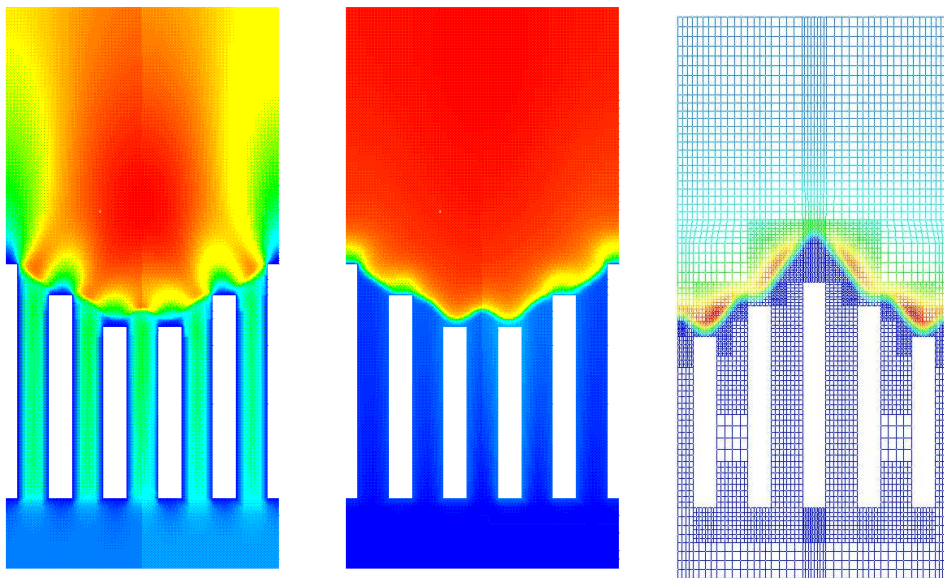


Figure 39: Results of the methane burner simulation: velocity and temperature profiles (left and middle), finest mesh with 5,000 cells (right); from Braack [17].

## References

- [1] M. Ainsworth and J. T. Oden (1997), *A posteriori error estimation in finite element analysis*, Comput. Meth. Appl. Mech. Engrg., 142, pp. 1-88.
- [2] T. Apel and M. Dobrowolski (1992), *Anisotropic interpolation with applications to the finite element method*, Computing, 47, pp. 277–293.
- [3] T. Apel (1999), *Anisotropic Finite Elements: Local Estimates and Applications*, Habilitation Thesis, Preprint 99-03, SFB 393, University of Magdeburg.
- [4] E. Backes (1997), *Gewichtete a posteriori Fehleranalyse bei der adaptiven Finite-Elemente-Methode: Ein Vergleich zwischen Residuen- und Bank-Weiser-Schätzer*, Diploma Thesis, Institute of Applied Mathematics, University of Heidelberg.
- [5] W. Bangerth and G. Kanschat (1999), *deal.II Homepage*, Technical Reference, Release 1.0, SFB 359, University of Heidelberg, <http://gaia.iwr.uni-heidelberg.de/~deal/>.
- [6] R. E. Bank, B. Weiser, and H. Yserentant (1990), *A class of iterative methods for solving saddle point problems*, Numer. Math., 56, 645–666.
- [7] R. Becker (1995), *An Adaptive Finite Element Method for the Incompressible Navier-Stokes Equations on Time-Dependent Domains*, Doctor Thesis, Preprint 95-44, SFB 359, Nov. 1995, University of Heidelberg.
- [8] R. Becker (1998), *An adaptive finite element method for the Stokes equations including control of the iteration error*, ENUMATH'95, Paris, Sept. 18-22, 1995, in: Proc. ENUMATH'97 (H. G. Bock, et al., eds.), pp. 609-620, World Scientific Publisher, Singapore.
- [9] R. Becker (1998): *Weighted error estimators for finite element approximations of the incompressible Navier-Stokes equations*, Preprint 98-20, SFB 359, University of Heidelberg, submitted for publication.
- [10] R. Becker, M. Braack, R. Rannacher, and C. Waguet (1998), *Fast and reliable solution of the Navier-Stokes equations including chemistry*, Proc. Conf. Applied Mathematics for Industrial Flow Problems (AMIF), San Feliu de Guixols (Spain), Oct. 1-3. 1998, Preprint 99-03 (SFB 359), University of Heidelberg, January 1999, to appear in Computing and Visualization in Sciences.
- [11] R. Becker, C. Johnson, and R. Rannacher (1995), *Adaptive error control for multigrid finite element methods*, Computing, 55, pp. 271-288.
- [12] R. Becker and R. Rannacher (1995), *Weighted a posteriori error control in FE methods*, ENUMATH'95, Paris, Sept. 18-22, 1995, Proc. ENUMATH'97 (H. G. Bock, et al., eds.), pp. 621–637, World Scientific Publishers, Singapore.

- [13] R. Becker and R. Rannacher (1994), *Finite element solution of the incompressible Navier-Stokes equations on anisotropically refined meshes*, Proc. Workshop “Fast Solvers for Flow Problems”, Kiel, Jan. 14-16, 1994 (W. Hackbusch and G. Wittum, eds.), pp. 52–61, NNFM, Vol. 49, Vieweg, Braunschweig.
- [14] R. Becker and R. Rannacher (1996), *A feed-back approach to error control in finite element methods: Basic analysis and examples*, East-West J. Numer. Math., 4, pp. 237-264.
- [15] H. Blum (1990), *Asymptotic Error Expansion and Defect Correction in the Finite Element Method*, Habilitation Thesis, University of Heidelberg.
- [16] M. Boman (1995), *A Model Study of Hydrodynamic Stability*, Masters Thesis, Chalmers University of Technology, Gothenburg, Sweden.
- [17] M. Braack (1998), *An Adaptive Finite Element Method for Reactive Flow Problems*, Doctor Thesis, Institute of Applied Mathematics, University of Heidelberg.
- [18] M. Braack and R. Rannacher (1999), *Adaptive finite element methods for low-Mach-number flows with chemical reactions*, Lecture Series 1999-03, 30th Computational Fluid Dynamics, (H. Deconinck, ed.), von Karman Institute for Fluid Dynamics, Belgium.
- [19] S. C. Brenner and R. L. Scott (1994), *The Mathematical Theory of Finite Element Methods*, Springer, Berlin-Heidelberg-New York.
- [20] F. Brezzi and M. Fortin (1991), *Mixed and Hybrid Finite Element Methods*, Springer, Berlin-Heidelberg-New York.
- [21] F. Brezzi and J. Pitkäranta (1984), *On the stabilization of finite element approximations of the Stokes equations*, Proc. Workshop Efficient Solution of Elliptic Systems (W. Hackbusch, ed.), Vieweg, Braunschweig.
- [22] M. O. Bristeau, R. Glowinski, and J. Periaux (1987), *Numerical methods for the Navier-Stokes equations: Applications to the simulation of compressible and incompressible viscous flows*, Comput. Phys. Reports, 6, pp. 73–187.
- [23] C. M. Chen and V. Thomée (1985), *The lumped mass finite element method for a parabolic problem*, J. Austral. Math. Soc., Ser. B, 26, pp. 329-354.
- [24] A. J. Chorin (1968), *Numerical solution of the Navier-Stokes equations*, Math. Comp., 22, pp. 745-762.
- [25] W. E and J. P. Liu (1998), *Projection method I: Convergence and numerical boundary layers*, SIAM J. Num. Anal., to appear.
- [26] K. Eriksson, D. Estep, P. Hansbo, and C. Johnson (1995), *Introduction to adaptive methods for differential equations*, Acta Numerica 1995 (A. Iserles, ed.), pp. 105-158, Cambridge University Press.

- [27] M. Feistauer (1993), *Mathematical Methods in Fluid Dynamics*, Longman Scientific&Technical, England.
- [28] M. L. M. Giles, M. Larson, and E. Süli (1998), *Adaptive error control for finite element approximations of the lift and drag coefficients in viscous flow*, SIAM J. Numer. Anal., to appear.
- [29] V. Girault and P.-A. Raviart (1986), *Finite Element Methods for the Navier-Stokes Equations*, Springer, Heidelberg.
- [30] R. Glowinski (1985), *Viscous flow simulations by finite element methods and related numerical techniques*, in Progress in Supercomputing in Computational Fluid Dynamics (E.M. Murman and S.S. Abarbanel, eds.), pp. 173-210, Birkhäuser, Boston.
- [31] R. Glowinski and J. Periaux (1987), *Numerical methods for nonlinear problems in fluid dynamics*, Proc. Int. Seminar on Scientific Supercomputers, Paris, North-Holland, Amsterdam.
- [32] P. M. Gresho (1990), *On the theory of semi-implicit projection methods for viscous incompressible flow and its implementation via a finite element method that also introduces a nearly consistent mass matrix, Part 1: Theory*, Int. J. Numer. Meth. Fluids, 11, pp. 587–620.
- [33] P. M. Gresho and S. T. Chan (1990), *On the theory of semi-implicit projection methods for viscous incompressible flow and its implementation via a finite element method that also introduces a nearly consistent mass matrix, Part 2: Implementation*, Int. J. Numer. Meth. Fluids, 11, pp. 621–660.
- [34] P. M. Gresho (1991), *Incompressible fluid dynamics: Some fundamental formulation issues*, Ann. Rev. Fluid Mech., 23, pp. 413-453.
- [35] P. M. Gresho and R. L. Sani (1998), *Incompressible Flow and the Finite Element Method*, John Wiley, Chichester.
- [36] W. Hackbusch (1985), *Multi-Grid Methods and Applications*, Springer, Heidelberg-Berlin.
- [37] P. Hanbo and C. Johnson (1995), *Streamline diffusion finite element methods for fluid flow*, in Finite Element Methods for Compressible and Incompressible Flow, Selected Topics from Previous VKI Lecture Series, (H. Deconinck, ed.), von Karman Institute for Fluid Dynamics, Belgium.
- [38] J. Harig (1991), *Eine robuste und effiziente Finite-Elemente Methode zur Lösung der inkompressiblen 3D-Navier-Stokes Gleichungen auf Vektorrechnern*, Doctor Thesis, Institute of Applied Mathematics, University of Heidelberg.
- [39] R. Hartmann (1998), *A posteriori Fehlerschätzung und adaptive Schrittweiten- und Ortsgittersteuerung bei Galerkin-Verfahren für die Wärmeleitungsgleichung*, Diploma Thesis, Institute of Applied Mathematics, University of Heidelberg.



- [40] J. G. Heywood (1980), *The Navier-Stokes equations: On the existence, regularity and decay of solutions*, Indiana Univ. Math. J., 29, pp. 639-681.
- [41] J. G. Heywood and R. Rannacher (1982), *Finite element approximation of the nonstationary Navier-Stokes Problem. I. Regularity of solutions and second order error estimates for spatial discretization*, SIAM J. Numer. Anal., 19, pp. 275-311.
- [42] J. G. Heywood and R. Rannacher (1986), *Finite element approximation of the nonstationary Navier-Stokes Problem. II. Stability of solutions and error estimates uniform in time*, SIAM J. Numer. Anal., 23, pp. 750-777.
- [43] J. G. Heywood and R. Rannacher (1988), *Finite element approximation of the nonstationary Navier-Stokes Problem. III. Smoothing property and higher order error estimates for spatial discretization*, SIAM J. Numer. Anal., 25, pp. 489-512.
- [44] J. G. Heywood and R. Rannacher (1990), *Finite element approximation of the nonstationary Navier-Stokes Problem. IV. Error analysis for second-order time discretization*, SIAM J. Numer. Anal., 27, pp. 353-384.
- [45] J. G. Heywood and R. Rannacher (1986), *An analysis of stability concepts for the Navier-Stokes equations*, J. Reine Angew. Math., 372, pp. 1-33.
- [46] J. G. Heywood, R. Rannacher, and S. Turek (1992), *Artificial boundaries and flux and pressure conditions for the incompressible Navier-Stokes equations*, Int. J. Numer. Math. Fluids, 22, pp. 325-352.
- [47] P. Houston, R. Rannacher, and E. Süli (1999), *A posteriori error analysis for stabilised finite element approximation of transport problems*, Report No. 99/04, Oxford University Computing Laboratory, Oxford, England OX1 3QD, submitted for publication.
- [48] T. J. R. Hughes and A. N. Brooks (1982), *Streamline upwind/Petrov-Galerkin formulations for convection dominated flows with particular emphasis on the incompressible Navier-Stokes equation*, Comput. Meth. Appl. Mech. Engrg., 32, pp. 199-259.
- [49] T. J. R. Hughes, L. P. Franca, and M. Balestra (1986), *A new finite element formulation for computational fluid mechanics: V. Circumventing the Babuska-Brezzi condition: A stable Petrov-Galerkin formulation of the Stokes problem accommodating equal order interpolation*, Comput. Meth. Appl. Mech. Engrg., 59, pp. 85-99.
- [50] C. Johnson (1987), *Numerical Solution of Partial Differential Equations by the Finite Element Method*, Cambridge University Press, Cambridge-Lund.
- [51] C. Johnson (1989), *The streamline diffusion finite element method for compressible and incompressible fluid flow*, Proc. Finite Element Method in Fluids VII, Huntsville.

- [52] C. Johnson (1993), *A new paradigm for adaptive finite element methods*, Proc. MAFELAP'93 Conf., Brunel Univ., Uxbridge, UK, John Wiley, Chichester.
- [53] C. Johnson, S. Larsson, V. Thomée, and L. B. Wahlbin (1987), *Error estimates for spatially discrete approximations of semilinear parabolic equations with nonsmooth initial data*, Math. Comp., 49, pp. 331-357.
- [54] C. Johnson, R. Rannacher, and M. Boman (1995), *Numerics and hydrodynamic stability: Towards error control in CFD*, SIAM J. Numer. Anal., 32, pp. 1058-1079.
- [55] C. Johnson, R. Rannacher, and M. Boman (1995), *On transition to turbulence and error control in CFD*, Preprint 95-06, SFB 359, University of Heidelberg.
- [56] C. Johnson and R. Rannacher (1994), *On error control in CFD*, Proc. Int. Workshop "Numerical Methods for the Navier-Stokes Equations", Heidelberg, Oct. 25-28, 1993 (F.-K. Hebeker, R. Rannacher, and G. Wittum, eds.), pp. 25-28, NNFM, Vol. 47, Vieweg, Braunschweig.
- [57] S. Kracmar and J. Neustupa (1994), *Modelling of flows of a viscous incompressible fluid through a channel by means of variational inequalities*, Z. Angew. Math. Mech., 74, pp. T637 - T639.
- [58] O. Ladyshenskaya (1969), *The Mathematical Theory of Viscous Incompressible Flow*, Gordon and Breach, London.
- [59] M. Landahl (1980), *A note on an algebraic instability of viscous parallel shear flows*, J. Fluid Mech., 98, p. 243.
- [60] P.-L. Lions (1998), *Mathematical Topics in Fluid Mechanics, Vol. 2, Compressible Models*, Clarendon Press, Oxford.
- [61] M. Luskin and R. Rannacher (1982), *On the smoothing property of the Crank-Nicolson scheme*, Appl. Anal., 14, pp. 117-135.
- [62] M. Luskin and R. Rannacher (1981), *On the smoothing property of the Galerkin method for parabolic equations*, SIAM J. Numer. Anal., 19, pp. 93-113.
- [63] K. W. Morton (1996), *Numerical Solution of Convection-Diffusion Problems*, Applied Mathematics and Mathematical Computation, Vol. 12, Chapman&Hall.
- [64] S. Müller-Urbaniak (1993), *Eine Analyse des Zwischenschritt- $\theta$ -Verfahrens zur Lösung der instationären Navier-Stokes-Gleichungen*, Doctor Thesis, Preprint 94-01, SFB 359, University of Heidelberg.
- [65] S. Müller, A. Prohl, R. Rannacher, and S. Turek (1994), *Implicit time-discretization of the nonstationary incompressible Navier-Stokes equations*, Proc. Workshop "Fast Solvers for Flow Problems", Kiel, Jan. 14-16, 1994 (W. Hackbusch and G. Wittum, eds.), pp. 175-191, NNFM, Vol. 49, Vieweg, Braunschweig.

- [66] H. Oswald (1999), *Lösung der instationären Navier-Stokes-Gleichungen auf Parallelrechnern*, Doctor Thesis, Institute of Applied Mathematics, University of Heidelberg.
- [67] O. Pironneau (1982), *On the transport-diffusion algorithm and its applications to the Navier-Stokes equations*, Numer. Math., 38, pp. 309–332.
- [68] O. Pironneau (1983), *Finite Element Methods for Fluids*, John Wiley, Chichester.
- [69] A. Prohl (1995), *Projektions- und Quasi-Kompressibilitätsmethoden zur Lösung der inkompressiblen Navier-Stokes-Gleichungen*, Doctor Thesis, Institute of Applied Mathematics, University of Heidelberg.
- [70] A. Prohl (1998), *On Quasi-Compressibility Methods and Projection Methods for Solving the Incompressible Navier-Stokes Equations*, Teubner, Stuttgart.
- [71] R. Rannacher (1984), *Finite element solution of diffusion problems with irregular data*, Numer. Math., 43, pp. 309–327.
- [72] R. Rannacher (1998), *Numerical analysis of nonstationary fluid flow (a survey)*, in “Applications of Mathematics in Industry and Technology” (V.C.Boffi and H.Neunzert, eds.), pp. 34–53, B.G. Teubner, Stuttgart.
- [73] R. Rannacher (1992), *On Chorin’s projection method for the incompressible Navier-Stokes equations*, Proc. Oberwolfach Conf. Navier-Stokes Equations: Theory and Numerical Methods, September 1991 (J.G. Heywood, et al., eds.), pp. 167–183, LNM, Vol. 1530, Springer, Heidelberg.
- [74] R. Rannacher (1993), *On the numerical solution of the incompressible Navier-Stokes equations*, Survey lecture at the annual GAMM-Conference 1992, Leipzig, March 24–27, Z. Angew. Math. Mech., 73, pp. 203–216.
- [75] R. Rannacher (1998), *A posteriori error estimation in least-squares stabilized finite element schemes*, Comput. Meth. Appl. Mech. Engrg., 166, pp. 99–114.
- [76] R. Rannacher (1999), *Error control in finite element computations*, Proc. NATO-Summer School “Error Control and Adaptivity in Scientific Computing” Antalya (Turkey), Aug. 1998 (H. Bulgak and C. Zenger, eds.), pp. 247–278, NATO Science Series, Series C, Vol. 536, Kluwer, Dordrecht.
- [77] R. Rannacher and S. Turek (1992), *A simple nonconforming quadrilateral Stokes element*, Numer. Meth. Part. Diff. Equ., 8, pp. 97–111.
- [78] R. Rautmann (1983), *On optimal regularity of Navier-Stokes solutions at time  $t = 0$* , Math. Z., 184, pp. 141–149.
- [79] H. Reichert and G. Wittum (1993), *On the construction of robust smoothers for incompressible flow problems*, Proc. Workshop “Numerical Methods for the Navier-Stokes Equations”, Heidelberg, Oct. 25–28, 1993 (F.K. Hebeker, R. Rannacher, G. Wittum, eds.), pp. 207–216, Vieweg, Braunschweig.

- [80] M. D. Smooke (1991), *Numerical Modeling of Laminar Diffusion Flames*, Progress in Astronautics and Aeronautics, 135.
- [81] M. Schäfer and S. Turek (1996), *The benchmark problem “flow around a cylinder”*, Flow Simulation with High-Performance Computers (E.H. Hirschel, ed.), pp. 547–566, NNFM, Vol. 52, Vieweg, Braunschweig.
- [82] F. Schieweck (1992), *A parallel multigrid algorithm for solving the Navier-Stokes equations on a transputer system*, Impact Comput. Sci. Engrg., 5, pp. 345-378.
- [83] P. Schreiber and S. Turek (1993), *An efficient finite element solver for the non-stationary incompressible Navier-Stokes equations in two and three dimensions*, Proc. Workshop “Numerical Methods for the Navier-Stokes Equations”, Heidelberg, Oct. 25-28, 1993 (F.K. Hebeker, R. Rannacher, G. Wittum, eds.), pp. 133-144, Vieweg, Braunschweig.
- [84] J. Shen (1992), *On error estimates of projection methods for the Navier-Stokes Equations: First order schemes*, SIAM J. Numer. Anal., 29, pp. 57-77.
- [85] J. Shen (1994), *Remarks on the pressure error estimates for the projection methods*, Numer. Math., 67, pp. 513-520.
- [86] J. Shen (1996), *On error estimates of the projection methods for the Navier-Stokes Equations: 2nd order schemes*, Math. Comp., 65, pp. 1039–1065.
- [87] G. Strang (1968), *On the construction and comparison of difference schemes*, SIAM J. Numer. Anal., 5, 506-517.
- [88] R. Temam (1987), *Navier-Stokes Equations. Theory and numerical analysis*, 2nd ed., North Holland, Amsterdam.
- [89] V. Thomée (1997), *Galerkin Finite Element Methods for Parabolic Problems*, Springer, Berlin-Heidelberg-New York.
- [90] L. Tobiska and F. Schieweck (1989), *A nonconforming finite element method of up-stream type applied to the stationary Navier-Stokes equation*, M<sup>2</sup>AN, 23, pp. 627-647.
- [91] L. Tobiska and R. Verfürth (1996), *Analysis of a streamline diffusion finite element method for the Stokes and Navier-Stokes equations*, SIAM J. Numer. Anal., 33, pp. 107–127.
- [92] L. N. Trefethen, A. E. Trefethen, S. C. Reddy, and T. A. Driscoll (1992), *A new direction in hydrodynamical stability: Beyond eigenvalues*, Tech. Report CTC92TR115 12/92, Cornell Theory Center, Cornell University.
- [93] S. Turek (1994), *Tools for simulating nonstationary incompressible flow via discretely divergence-free finite element models*, Int. J. Numer. Meth. Fluids, 18, pp. 71-105.

- [94] S. Turek (1996), *A comparative study of some time-stepping techniques for the incompressible Navier-Stokes equations*, Int. J. Numer. Meth. Fluids, 22, pp. 987–1011.
- [95] S. Turek (1997), *On discrete projection methods for the incompressible Navier-Stokes equations: An algorithmic approach*, Comput. Methods Appl. Mech. Engrg., 143, pp. 271–288.
- [96] S. Turek (1998), *FEATFLOW: Finite Element Software for the Incompressible Navier-Stokes Equations*, User Manual, Release 1.3, SFB 359, University of Heidelberg, URL: <http://gaia.iwr.uni-heidelberg.de/~featflow/>.
- [97] S. Turek (1999), *Efficient Solvers for Incompressible Flow Problems*, NNCSE, Vol. 6, Springer, Berlin-Heidelberg-New York.
- [98] S. Turek, et al. (1999), *Virtual Album of Fluid Motion*, Preprint, Institute of Applied Mathematics, University of Heidelberg, in preparation.
- [99] M. Van Dyke (1982), *An Album of Fluid Motion*, The Parabolic Press, Stanford.
- [100] J. Van Kan (1986), *A second-order accurate pressure-correction scheme for viscous incompressible flow*, J. Sci. Stat. Comp., 7, pp. 870-891.
- [101] S. P. Vanka (1986), *Block-implicit multigrid solution of Navier-Stokes equations in primitive variables*, J. Comp. Phys., 65, pp. 138–158.
- [102] R. Verfürth (1996), *A Review of A Posteriori Error Estimation and Adaptive Mesh-Refinement Techniques*, Wiley/Teubner, New York-Stuttgart, 1996.
- [103] C. Waguët (1999) *Adaptive Finite Element Computation of Chemical Flow Reactors*, Doctor Thesis, SFB 359, University of Heidelberg, in preparation.
- [104] P. Wesseling (1992), *An Introduction to Multigrid Methods*, J. Wiley, Chichester.
- [105] G. Wittum (1990), *The use of fast solvers in computational fluid dynamics*, in Proc. Eighth GAMM-Conference on Numerical Methods in Fluid Mechanics (P. Wesseling, ed.), pp. 574–581. LNFM, Vol. 29, Vieweg, Braunschweig.
- [106] O. C. Zienkiewicz and J. Z. Zhu (1987), *A simple error estimator and adaptive procedure for practical engineering analysis*, Int. J. Numer. Meth. Engrg., 24, pp. 337-357.
- [107] Guohui Zhou (1997), *How accurate is the streamline diffusion finite element method?*, Math. Comp., 66, pp. 31–44.