COMPUTATIONAL METHODS IN APPLIED MATHEMATICS, Vol. 6 (2006), No. 3, pp. 239–263

TRANSPORT-STABILIZED SEMIDISCRETIZATIONS OF THE INCOMPRESSIBLE NAVIER — STOKES EQUATIONS

L. ANGERMANN[1]

Abstract — Within the framework of finite element methods, the paper investigates a general approximation technique for the nonlinear convective term of Navier — Stokes equations. The approach is based on an upwind method of the finite volume type. It has been proved that the discrete convective term satisfies the well-known collection of sufficient conditions for convergence of the finite element solution. For a particular nonconforming scheme, the assumptions have been verified in detail and the estimate of the semidiscrete velocity error has been proved.

2000 Mathematics Subject Classification: 65N15, 65N30, 76D05.

Keywords: Navier — Stokes equations, nonconforming finite elements, upwind stabilization, finite volume methods.

1. Introduction

The system of incompressible Navier — Stokes equations is one of the most interesting and challenging models in computational fluid dynamics (CFD). A particular problem is the choice of stabilization approaches for the case of high Reynolds numbers.

The present paper focuses on this and describes a general approach to the design and analysis of discretization methods for the Navier — Stokes equations of viscous incompressible homogeneous fluids, where the stabilization effect is based on so-called FVM-based upwind methods.

For particular discretizations of the stationary system, a lot of work on error analysis (including numerical illustrations) has been done by Schieweck, Tobiska and co-workers [17–21]. Kanayama and Toshigami [12] have applied Ikeda's "partial upwind scheme E" [11] (see also [5]) to the Navier — Stokes equations and Miller and Wang [14] have described an exponentially fitted finite volume method for the streamline-vorticity formulation, but in both papers no analysis is given. In the papers by Feistauer and co-authors [4, 8–10], the so-called "combined method" has been investigated, where the finite volume method is applied to the convection terms and the resulting formulation is interpreted in a variational context as within the finite element method. However, these papers mainly treat the case of simplicial finite element partitions with finite volumes of the barycentric type.

For the present work, the papers [2] and [3] served as starting points. In the first of these papers, an attempt was made to extract the underlying principles of FVM-type discretizations for the stationary Navier — Stokes equations. In the second paper, an overview on certain problems in the full discretization of linear parabolic problems is given for the case where the semidiscretization in space is done by finite volume methods.

[1] *Technische Universität Clausthal, Institut für Mathematik*, Erzstraße 1, D–38678 Clausthal, Germany. E-mail: angermann@math.tu-clausthal.de

Here, some ideas of both papers are merged to get a result on the non-stationary Navier — Stokes equations. In this respect the proposed discretization can be viewed as a variant of the “hybrid” or “combined” finite volume–finite element method. In addition to [2], we will investigate in more detail a particular discretization method. This is based on the second part of the preprint [1], the first part of which has been published as [2].

The key point of the approach is the treatment of the convective term which is usually considered in a variational context as a trilinear form. In fact, a detailed study of Sect. IV.2 in [17], where a complete convergence analysis is given for the spatial discretization of the stationary Navier — Stokes equations by means of the Crouzeix and Raviart element, shows that there are three distinguished properties of the discrete trilinear form that guarantee (first-order) convergence, provided the family of finite element spaces satisfies a discrete inf-sup condition. These properties are (a precise formulation will be given in the next section):

(i) semidefiniteness,

(ii) Lipschitz-continuity,

(iii) (linear) consistency.

The basic principle of the discretization method for the trilinear form such that these properties can be fulfilled stems from finite volume methods which have been successfully applied in many situations, especially when it is important to have a discrete conservation law or a discrete maximum principle. In contrast to many standard approaches for finite volume methods, where the design of the control volumes follows narrow rules, in our approach the control volumes can be chosen relatively free. In particular, the correlation to other partitions of the domain is not very strong.

The paper is subdivided into four parts. The introductory section is followed by a technical section where the basic notation is described and both the weak and semidiscrete problems are formulated. In the third section, we discuss the discretization of the convective term. The main part illustrates the basic aspects of the theory by demonstrating the application to a finite element method due to Schieweck [19], where the particular treatment of the trilinear form slightly differs from Schieweck's originally proposed one. We will demonstrate an estimate of the semidiscrete velocity error measured in a discrete L_2-norm without use of any linearized stability theory. As long as the numerical solution satisfies a certain smallness assumption, the stationary pendant of which is widely used (cf. [17, Sect. IV.2]), it will be shown that the constant in the error estimate is time-independent and of order $\mathcal{O}(\varepsilon^{-1/2}) = \mathcal{O}(\sqrt{Re})$ but not $\mathcal{O}(\exp(\varepsilon^{-1})) = \mathcal{O}(\exp(Re))$.

2. Notation and preliminaries

2.1. Formulation of the problem. Let $\varepsilon > 0$ be a real number, $\Omega \subset \mathbf{R}^d$ with $d = 2$ or $d = 3$ be a bounded, polyhedral domain with Lipschitz-continuous boundary, $t_\infty > 0$ and $f : (0, t_\infty) \times \Omega \to \mathbf{R}^d$, $u_0 : \Omega \to \mathbf{R}^d$ be given vector fields. The following nonlinear system of partial differential equations with respect to the $d+1$ variables $u = (u^1, \dots, u^d)^\top : (0, t_\infty) \times \Omega \to \mathbf{R}^d$, $p : (0, t_\infty) \times \Omega \to \mathbf{R}$ is considered:

$$\begin{aligned} &\partial_t u - \varepsilon \Delta u + (u \cdot \nabla) u + \nabla p = f \quad \text{in} \quad (0, t_\infty) \times \Omega, \\ &\nabla \cdot u = 0 \quad \text{in} \quad (0, t_\infty) \times \Omega, \\ &u = 0 \quad \text{on} \quad (0, t_\infty) \times \partial\Omega, \\ &u = u_0 \quad \text{on} \quad \{0\} \times \Omega \end{aligned} \tag{2.1}$$

(Navier — Stokes equations of viscous incompressible homogeneous fluids with homogeneous nonslip boundary condition).

In order to give a weak formulation of this problem, we introduce the function spaces $V := \overset{\circ}{W}{}^1_2(\Omega)^d$, $Q := L_{2,0}(\Omega) := \{q \in L_2(\Omega) : (q,1) = 0\}$, $W := \{v \in L_2(\Omega)^d : (q, \nabla \cdot v) = 0\ \forall q \in Q\}$ and define a trilinear form $n : V^3 \to \mathbf{R}$ by

$$n(w,u,v) := ((w \cdot \nabla)u, v), \tag{2.2}$$

where $(\cdot,\cdot)$ denotes the $L_2(\Omega)$- or $L_2(\Omega)^d$-inner product.

If the symbol ∇ is applied to a vector field, say $v \in V$, then, as usual, it will be understood as a tensor of order two with elements $(\nabla v)_{jl} := \partial_j v^l$, $j, l = 1, \ldots, d$, where ∂_j denotes differentiation with respect to the j-th spatial variable.

Finally, for two vector fields $v, w \in V$, the bilinear form $(\nabla v, \nabla w)$ is defined by $(\nabla v, \nabla w) := \sum_{l=1}^d (\nabla v^l, \nabla w^l)$. Then, given $f \in L_2((0,t_\infty), V^*)$ and $u_0 \in W$, the corresponding weak formulation of (2.1) reads as follows:

Find $(u,p) \in L_2((0,t_\infty), V) \times L_2((0,t_\infty), Q)$ such that

$$(\partial_t u, v) + \varepsilon(\nabla u, \nabla v) + n(u,u,v) - (p, \nabla \cdot v) + (q, \nabla \cdot u) = (f, v) \quad \forall (v,q) \in V \times Q,$$

$$u = u_0 \quad \text{on} \quad \{0\} \times \Omega, \tag{2.3}$$

where the variational equation holds, on $(0,t_\infty)$, in the sense of distributions.

Now let us suppose that there are given two finite-dimensional spaces $V_h \subset L_2(\Omega)^d$, $Q_h \subset L_2(\Omega)$ approximating, in a certain sense, the spaces V, Q. In general, these discrete spaces need not be subspaces of V and Q, respectively.

Typically, they consist of piecewise polynomial functions with respect to certain partitions of the domain Ω. While it is not difficult to replace the forms

$$(\nabla u, \nabla v) \quad \text{and} \quad (p, \nabla \cdot v)$$

by their “broken” counterparts on the underlying partitions $\mathcal{T}_h$ of Ω

$$(\nabla u, \nabla v)_h := \sum_{T \in \mathcal{T}_h} (\nabla u, \nabla v)_T \quad \text{and} \quad (p, \nabla \cdot v)_h := \sum_{T \in \mathcal{T}_h} (p, \nabla \cdot v)_T$$

and to analyze the resulting properties, the trilinear form n has to be defined in a more careful way for stability reasons. In the above formulas, the subscript T indicates the restriction of the integration domain on the subset $T \subset \Omega$.

It was pointed out in [17, Sect. IV.2] that if the finite element spaces $V_h \times Q_h$ are stable (i.e., they satisfy a discrete inf-sup condition), then essentially the following three properties of the discrete form $n_h : V_h^3 \to \mathbf{R}$ are sufficient conditions for establishing convergence of the numerical method for the stationary incompressible Navier — Stokes equations:

semidefiniteness: $n_h(w_h, v_h, v_h) \geqslant 0$,

Lipschitz-continuity: $|n_h(w_h, u_h, v_h) - n_h(z_h, u_h, v_h)| \leqslant C\|w_h - z_h\|_h \|u_h\|_h \|v_h\|_h$,

consistency: $|n(w, u, v_h) - n_h(I_h w, I_h u, v_h)| \leqslant Ch\|w\|_{2,2,\Omega}\|u\|_{2,2,\Omega}\|v_h\|_h$,

$u_h, v_h, w_h, z_h \in V_h$, $u, w \in W_2^2(\Omega)^d \bigcap V$,

where $\|\cdot\|_h$ is a norm on V_h and $I_h : V \to V_h$ denotes some interpolation operator.

Here we will show that these conditions allow to formulate a similar result for a non-stationary situation.

We consider the following semidiscrete formulation, where $(\cdot,\cdot)_l$ denotes a discrete $L_2(\Omega)^d$-inner product that is different from $(\cdot,\cdot)_h$ in general: for the given approximation $u_{0h} \in V_h$ to u_0, find $(u_h, p_h) \in L_2((0,t_\infty), V_h \times Q_h)$ with $\partial_t u_h \in L_2((0,t_\infty), L_2(\Omega)^d)$ such that

$$(\partial_t u_h, v_h)_l + \varepsilon(\nabla u_h, \nabla v_h)_h + n_h(u_h, u_h, v_h) - (p_h, \nabla\cdot v_h)_h + (q_h, \nabla\cdot u_h)_h = (f, v_h)_l \quad \forall (v_h, q_h) \in V_h \times Q_h,$$

$$u_h = u_{0h} \quad \text{on} \quad \{0\}\times\Omega. \tag{2.4}$$

In order to give an overview on what follows we will sketch the typical steps of the proof of convergence of the semidiscrete solution u_h (provided it exists uniquely) to the weak solution u.

If the weak solution (u,p) of (2.3) additionally belongs to

$$L_2((0,t_\infty), W_2^2(\Omega)^d) \times L_2((0,t_\infty), W_2^1(\Omega)), \tag{2.5}$$

the variational equation in (2.3), restricted to the test space $V \times \{0\}$, can be written as follows:

$$(\partial_t u, v) - \varepsilon(\Delta u, v) + n(u,u,v) + (\nabla p, v) = (f, v) \qquad \forall v \in V \tag{2.6}$$

i.e., by Sobolev's embedding theorem, the first equation of (2.1) is satisfied in space in the $L_2(\Omega)^d$-sense. Therefore, since $V_h \subset L_2(\Omega)^d$, equation (2.6) makes sense for test functions from V_h, too. After a simple manipulation, we arrive at the following equation:

$$(\partial_t u, v_h) + \varepsilon(\nabla u, \nabla v_h)_h + ((u\cdot\nabla)u, v_h) - (p, \nabla\cdot v_h)_h = (f, v_h) + \langle \tilde{f}_h, v_h\rangle \quad \forall v_h \in V_h, \tag{2.7}$$

where $\langle \tilde{f}_h, v_h\rangle := \varepsilon(\Delta u, v_h) + \varepsilon(\nabla u, \nabla v_h)_h - (\nabla p, v_h) - (p, \nabla\cdot v_h)_h$.

Remark 2.1. Rewriting the terms $(\Delta u, v_h)$ and $(\nabla p, v_h)$, where $v_h \in V_h$, in the element-by-element manner and integrating by parts locally on each element $T \in \mathfrak{T}_h$, the consistency error functional can be represented as follows:

$$\langle \tilde{f}_h, v_h\rangle := \sum_{T\in\mathfrak{T}_h}\Big[\varepsilon\Big(\frac{\partial u}{\partial\nu}, v_h\Big)_{\partial T} - (p, \nu\cdot v_h)_{\partial T}\Big],$$

where $(\partial u/\partial\nu, v_h)_{\partial T} := \sum_{l=1}^d (\nu\cdot\nabla u^l, v_h^l)_{\partial T}$.

Finally, for the following it will be convenient to use an element-by-element version $\tilde{n}_h$ of n:

$$\tilde{n}_h(w,u,v) := \sum_{T\in\mathfrak{T}_h} ((w\cdot\nabla)u, v)_T.$$

Then the above equation (2.7) reads as

$$(\partial_t u, v_h) + \varepsilon(\nabla u, \nabla v_h)_h + \tilde{n}_h(u,u,v_h) - (p, \nabla\cdot v_h)_h = (f, v_h) + \langle \tilde{f}_h, v_h\rangle \qquad \forall v_h \in V_h. \tag{2.8}$$

Restricting the semidiscrete formulation (2.4) to the test space $V_h \times \{0\}$ and subtracting the result from (2.8), we obtain

$$(\partial_t u, v_h) - (\partial_t u_h, v_h)_l + \varepsilon(\nabla(u-u_h), \nabla v_h)_h = n_h(u_h, u_h, v_h) - \tilde{n}_h(u,u,v_h) +$$

$$(p - p_h, \nabla\cdot v_h)_h + \ (f, v_h) + \langle \tilde{f}_h, v_h\rangle - (f, v_h)_l \quad \forall v_h \in V_h. \tag{2.9}$$

For arbitrary elements $v_h, w_h \in V_h$, $(\partial_t u, v_h) = (\partial_t u, v_h)_l + (\partial_t u, v_h) - (\partial_t u, v_h)_l = (\partial_t w_h, v_h)_l + (\partial_t(u - w_h), v_h)_l + (\partial_t u, v_h) - (\partial_t u, v_h)_l$ holds.

Thanks to $u - u_h = (w_h - u_h) + (u - w_h)$, we get from (2.9) that

$$(\partial_t(w_h - u_h), v_h)_l + \varepsilon(\nabla(w_h - u_h), \nabla v_h)_h = n_h(u_h, u_h, v_h) - \tilde{n}_h(u, u, v_h) + (p - p_h, \nabla \cdot v_h)_h - \\ (\partial_t(u - w_h), v_h)_l + (\partial_t u, v_h)_l - (\partial_t u, v_h) - \varepsilon(\nabla(u - w_h), \nabla v_h)_h + (f, v_h) + \langle \tilde{f}_h, v_h \rangle - (f, v_h)_l \quad \forall v_h \in V_h. \tag{2.10}$$

Furthermore, we have

$$n_h(u_h, u_h, v_h) = n_h(w_h, w_h, v_h) + n_h(u_h, u_h, v_h) - n_h(w_h, w_h, v_h) = \\ n_h(w_h, w_h, v_h) + n_h(u_h, u_h, v_h) - n_h(w_h, u_h, v_h) - n_h(w_h, w_h - u_h, v_h),$$

hence we finally arrive at the following velocity error equation:

$$\begin{aligned} &(\partial_t(w_h - u_h), v_h)_l + \varepsilon(\nabla(w_h - u_h), \nabla v_h)_h = \\ &n_h(w_h, w_h, v_h) - \tilde{n}_h(u, u, v_h) && (2.11) \\ &+ \ n_h(u_h, u_h, v_h) - n_h(w_h, u_h, v_h) && (2.12) \\ &- \ n_h(w_h, w_h - u_h, v_h) && (2.13) \\ &+ \ (p - p_h, \nabla \cdot v_h)_h && (2.14) \\ &- \ (\partial_t(u - w_h), v_h)_l && (2.15) \\ &+ \ (\partial_t u, v_h)_l - (\partial_t u, v_h) && (2.16) \\ &- \ \varepsilon(\nabla(u - w_h), \nabla v_h)_h && (2.17) \\ &+ \ (f, v_h) - (f, v_h)_l && (2.18) \\ &+ \ \langle \tilde{f}_h, v_h \rangle \qquad \forall v_h \in V_h. && (2.19) \end{aligned}$$

If we set $v_h := w_h - u_h$ with $w_h := I_h u$, we can apply a standard energy argument provided we are able to estimate the terms (2.11) – (2.19) in an appropriate manner. That is, in the subsequent sections we have to investigate the following aspects:

- definition of n_h and the three properties mentioned above,
- definition of the interpolation operator $I_h : V \to V_h$ and its properties,
- definition of the interpolation operator $J_h : Q \to Q_h$ and its properties,
- definition of the lumping operator $L_h : C(\overline{\Omega})^d + V_h \to L_\infty(\Omega)$ generating $(\cdot, \cdot)_l$ and its properties,
- consistency error caused by the use of the broken inner product $(\cdot, \cdot)_h$.

2.2. Geometrical definitions and relations. The discretization procedure is based on three different families of partitions of Ω. An element of the first family of (primary) partitions is denoted by $\mathcal{T}_h$ and is either a triangulation (i.e., it consists of d-simplices) or a block-partition (i.e., it consists of convex quadrilaterals ($d = 2$) or convex hexahedra ($d = 3$)). It is assumed that $\mathcal{T}_h$ is admissible in the usual sense, i.e., two elements of the partition are allowed to have in common either a vertex or a complete edge or, if $d = 3$, a complete face. Using the notation T for the elements of $\mathcal{T}_h$, the parameter h of the partition is defined as follows: If T has the diameter h_T, then $h := \max_{T \in \mathcal{T}_h} h_T$. Notice that this partition is related to the approximation u_h of the unknown u; in certain situations the partition for the discrete unknown p_h may differ from $\mathcal{T}_h$.

Next, on each partition $\mathcal{T}_h$ a (not necessarily conforming) finite element space is defined, whose elements are piecewise polynomials of maximal degree $l \in \mathbf{N} \bigcup \{0\}$. In particular, the polynomial space on T may be incomplete, especially for quadrilateral/hexahedral elements.

Given some finite element space, the corresponding set of functionals (global degrees of freedom) naturally splits into Langrangian functionals and others, where a Langrangian functional is defined via the point values of its argument. Therefore, considering these Langrangian functionals, a collection of (global) nodes, called Langrangian nodes, can be associated in a natural way. For example, Langrangian nodes may be the vertices of triangles or the barycenters of faces of hexahedrons, where the above admissibility assumption allows to identify nodes with the same geometrical position. This collection of nodes can be subdivided into the class of nodes lying on element boundaries and the class of nodes belonging to the interior of some element.

Let $\overline{\Lambda}_g$ denote the set of indices of all Langrangian nodes from the first class. The subset $\Lambda_g \subset \overline{\Lambda}_g$ contains, by definition, the indices of interior (w.r.t. Ω) nodes only. Finally, declare $\partial\Lambda_g := \overline{\Lambda}_g \setminus \Lambda_g$ and let $\Lambda_{gT} \subset \overline{\Lambda}_g$ contain the (global) indices of the nodes belonging to the element T. Due to the boundary conditions in (2.1), the above finite element space is restricted to elements satisfying a discrete boundary condition, i.e., we set

$$S_{hl} := \{v_h \,:\, (v_h|_T \in \mathcal{P}_l(T)\ \forall T \in \mathcal{T}_h) \wedge (v_h(x_i) = 0\ \forall i \in \partial\Lambda_g)\}\,.$$

The distance between two nodes x_i, x_j $(i, j \in \overline{\Lambda}_g)$ is denoted by d_{ij}.

Now, an important observation is that the index set $\overline{\Lambda}_g$ can be decomposed into two disjoint subsets $\overline{\Lambda}$, $\overline{\Lambda}^{\Delta}$ such that $\overline{\Lambda}_g = \overline{\Lambda} \bigcup \overline{\Lambda}^{\Delta}$ and $\overline{\Lambda} \bigcap \overline{\Lambda}^{\Delta} = \varnothing$. Such a decomposition can be generated, for example, by a hierarchical decomposition of the finite element space S_{hl} into a "lower degree part" and its linear complement. Then $\overline{\Lambda}$ corresponds to the nodes of the first part and $\overline{\Lambda}^{\Delta}$ to the nodes of the complement. Obviously, this decomposition induces similar decompositions of Λ_g, $\partial\Lambda_g$ and Λ_{gT}, respectively. If S_{hl} is a space of elements of low degree, then it is allowed that the decomposition is trivial, i.e., the complement may be the trivial space consisting of the zero element only. In this case, $\overline{\Lambda}^{\Delta}$ is empty by definition.

To describe the discretization of the trilinear form n, a further family of partitions of Ω is needed. The element $\mathcal{T}_h^*$ of the second family consists of subdomains $\Omega_i \subset \Omega$, whose boundary part $\Omega \bigcap \partial\Omega_i$ is a union of subsets of $(d-1)$-dimensional hyperplanes.

The incidence relation between these two partitions is defined with the help of the nodes of $\mathcal{T}_h$, i.e., each Ω_i should correspond to one node x_i and vice versa.

So we assume that a collection of points $\overline{\mathcal{N}} := \{x_i\}_{i\in\overline{\Lambda}} \subset \overline{\Omega}$, where $\overline{\Lambda} \subset \mathbf{N}$ is a finite index set, is given. Furthermore we assume that $\overline{\Lambda}$ is split into two disjoint subsets $\Lambda \subset \overline{\Lambda}$, $\partial\Lambda := \overline{\Lambda} \setminus \Lambda$ such that $\mathcal{N} := \{x_i\}_{i\in\Lambda} \subset \Omega$. Then we also have the decomposition $\overline{\mathcal{N}} = \mathcal{N} \bigcup \partial\mathcal{N}$ with $\partial\mathcal{N} := \{x_i\}_{i\in\partial\Lambda}$.

Remark 2.2. In this setting, the situation that $\partial\mathcal{N} \bigcap \Omega \neq \varnothing$ is not excluded. In some applications, $\partial\mathcal{N}$ may consist of points close to the boundary $\partial\Omega$ as well as of points lying at the boundary.

The set $\mathcal{T}_h^* = \{\Omega_i\}_{i\in\overline{\Lambda}}$ of control volumes Ω_i is assumed to satisfy the following properties:

(A1) $\mathcal{T}_h^*$ is a partition of Ω.

(A2) (i) $\forall i \in \Lambda : x_i \in \Omega_i$,

(ii) $\forall i \in \partial\Lambda : x_i \in \Omega_i \bigcup (\partial\Omega \bigcap \partial\Omega_i)$.

(A3) $\forall i \in \overline{\Lambda} : \Omega \bigcap \partial\Omega_i$ is a union of a finite number of subsets of $(d-1)$-dimensional hyperplanes.

For all indices $i, j \in \overline{\Lambda}$, $j \neq i$, let $\Gamma_{ij} := \overline{\Omega}_i \bigcap \overline{\Omega}_j$, $m_{ij} := \operatorname{meas}_{d-1}(\Gamma_{ij})$ and $x_{ij} := (x_i + x_j)/2$. The distance between two nodes x_i, x_j is denoted by d_{ij}. Then it makes sense to introduce the index set

$$\Lambda_i := \left\{ j \in \overline{\Lambda} \setminus \{i\} \; : \; m_{ij} > 0 \right\}, \quad i \in \overline{\Lambda}.$$

As a consequence of these definitions, for $i \in \Lambda$ the following representation of $\partial\Omega_i$ is valid: $\partial\Omega_i = \bigcup_{j \in \Lambda_i} \Gamma_{ij}$. Moreover, there are obvious symmetry relations

$$d_{ji} = d_{ij}, \quad \Gamma_{ji} = \Gamma_{ij}, \quad m_{ji} = m_{ij}. \tag{2.20}$$

If, in addition, the primary partition $\mathcal{T}_h = \{T_i\}_{i=1}^{n_h}$, $n_h \in \mathbf{N}$, has the node set $\overline{\Lambda}_g = \overline{\mathcal{N}}$, then a natural correlation between the two partitions $\mathcal{T}_h$, $\mathcal{T}_h^*$ exists and some further assumptions are necessary. Before formulating these assumptions, some additional notation has to be introduced.

$\mathcal{E}_T$ is the set of all faces (i.e., $(d-1)$-dimensional hypersurfaces) $E \subset \partial T$ of the element T. $\Lambda_T \subset \overline{\Lambda}$ denotes the set of indices of the nodes of the element $T \in \mathcal{T}_h$. For a face $E \in \mathcal{E}_T$ of some element $T \in \mathcal{T}_h$, $\Lambda_E \subset \Lambda_T$ denotes the set of indices of the nodes of E. Finally, for $i \in \Lambda_T$, we set $\Omega_{i,T} := \Omega_i \bigcap T$ for the *element control volume* induced by Ω_i.

(A4) $\forall i \in \overline{\Lambda} : \Omega_i \subset \bigcup_{T \in \mathcal{T}_h : \Lambda_T \ni i} \overline{T}$.

(A5) $\forall T \in \mathcal{T}_h \; \forall i \in \Lambda_T : \Omega_{i,T}$ is an open, simply connected, strongly Lipschitz set.

(A6) $\forall T \in \mathcal{T}_h \; \forall i \in \Lambda_T$: The set $\partial(\Omega_{i,T}) \bigcap \partial T \setminus \bigcup_{E \in \mathcal{E}_T : \Lambda_E \ni i} \overline{E}$ consists of at most one point[1] of ∂T.

(A7) $\forall T \in \mathcal{T}_h \; \forall i \in \Lambda_T \; \forall j \in \Lambda_T \bigcap \Lambda_i : x_{ij} \in \partial\Omega_{i,T}$.

Obviously, for $T \in \mathcal{T}_h$, $i \in \Lambda_T$ and $j \in \Lambda_T \bigcap \Lambda_i$, the boundary parts Γ_{ij} can be structured in a finer way: $\Gamma_{ij}^T := \Gamma_{ij} \bigcap T$. Analogously, $m_{ij}^T := \operatorname{meas}_{d-1}(\Gamma_{ij}^T)$.

(A8) There exists a third partition $\left\{\Omega_{ij}^T\right\}_{T \in \mathcal{T}_h, i \in \Lambda_T, j \in \Lambda_T \bigcap \Lambda_i : m_{ij}^T > 0}$ of Ω such that the subdomains Ω_{ij}^T have the following properties:

  (*i*) $\Omega_{ij}^T = \Omega_{ji}^T$,

  (*ii*) $\Gamma_{ij}^T \subset \overline{\Omega_{ij}^T}$, $x_i \in \overline{\Omega_{ij}^T}$, $x_j \in \overline{\Omega_{ij}^T}$.

  (*iii*) Each Ω_{ij}^T can be decomposed into a finite number l_{ij}^T of pairwise disjoint open d-simplices $\Omega_{ij}^{T,l}$ such that $\overline{\Omega}_{ij}^T = \bigcup_{l=1}^{l_{ij}^T} \overline{\Omega}_{ij}^{T,l}$ and, for any $l \in [1, l_{ij}^T]_N$, $\Omega_{ij}^{T,l}$ is the image of a fixed (reference) simplex $\hat{T}$ under a regular affine transformation, where the pre-image $\hat{\Gamma}$ of $\Gamma_{ij}^{T,l} := \Gamma_{ij} \bigcap \overline{\Omega}_{ij}^{T,l}$ does not depend on particular values of i, j, T, l.

  (*iv*) On each $\Gamma_{ij}^{T,l}$, the unit outer (w.r.t. Ω_i) normal $\nu_{ij}^{T,l}$ is constant.

The diameter of the simplex $\Omega_{ij}^{T,l}$ is denoted by $h_{ij}^{T,l}$. Furthermore, we set $m_{ij}^{T,l} := \operatorname{meas}_{d-1}(\Gamma_{ij}^{T,l})$ and $\Omega_{ij} := \operatorname{int}(\bigcup_{T \in \mathcal{T}_h : \Gamma_{ij}^T \neq \varnothing} \overline{\Omega}_{ij}^T)$.

Finally, in some cases certain regularity conditions are also needed.

[1] The (stronger) condition $\partial(\Omega_{i,T}) \bigcap \partial T \subset \bigcup_{E \in \mathcal{E}_T : \Lambda_E \ni i} \overline{E}$ is not satisfied for Voronoi diagrams on Friedrichs — Keller triangulations.

(A9) There exists a constant $C > 0$ such that for all $i, j \in \overline{\Lambda}$, $d_{ij}^5 \leqslant C m_{ij}$ holds.

(A10) There exists a constant $C > 0$ such that for all $i, j \in \overline{\Lambda}$,

$$d_{ij}\left(\max_{T\in\mathfrak{T}_h\,:\,m_{ij}^T>0}\ \max_{l\in[1,l_{ij}^T]_N} h_{ij}^{T,l}\right)^2 \leqslant C m_{ij}$$

holds.

(A11) There exists a constant $C > 0$ such that for all $i, j \in \overline{\Lambda},\ T \in \mathfrak{T}_h,\ l \in [1, l_{ij}^T]_N$,

$$m_{ij}^{T,l} d_{ij} \leqslant C \mathrm{meas}_d(\Omega_{ij}^{T,l})$$

holds.

(A12) There exists a constant $C > 0$ such that for all $i, j \in \overline{\Lambda},\ T \in \mathfrak{T}_h,\ l \in [1, l_{ij}^T]_N$,

$$(h_{ij}^{T,l})^4 \leqslant C \mathrm{meas}_d(\Omega_{ij}^{T,l})$$

holds.

Remark 2.3. Dimensional analysis of the quantities that appear in Assumptions (A9), (A10), (A12) easily shows that these conditions are not very restrictive.

3. Discretization of the trilinear form n

Using the decomposition

$$n(w, u, v) = \sum_{l=1}^{d}((w\cdot\nabla)u^l, v^l),$$

the description of the discretization can be reduced to the scalar case. So let $w \in \overset{\circ}{W}{}_2^1(\Omega)^d$ be such that $\nabla\cdot w = 0$ and define, for $u, v \in \overset{\circ}{W}{}_2^1(\Omega)$, the form

$$n_s(w, u, v) := ((w\cdot\nabla)u, v) = (w\cdot\nabla u, v).$$

The transport stabilization will be controlled by some *control function* $r : \mathbf{R} \to [0, 1]$ satisfying the following properties:

(P1) $r(z)$ is monotone for all z,

(P2) $\lim\limits_{z\to-\infty} r(z) = 0,\ \lim\limits_{z\to\infty} r(z) = 1$,

(P3) $1 + zr(z) \geqslant 0$ for all z,

(P4) $[1 - r(z) - r(-z)]z = 0$ for all z,

(P5) $[r(z) - 1/2]z \geqslant 0$ for all z,

(P6) $zr(z)$ is Lipschitz-continuous on the whole real axis.

The choice of exactly this control function dictates the upwind strategy of the numerical method. Typical examples of such control functions are:

$$r(z) = [\operatorname{sign} z + 1]/2 \quad \text{(full upwind scheme)},$$

$$r(z) = 1 - \frac{1}{z}\left[1 - \frac{z}{\exp z - 1}\right] \quad \text{(exponentially fitted scheme)}.$$

It is wellknown that the full upwind scheme introduces too much artificial diffusion. Although the exponentially fitted scheme should be preferred for theoretical reasons, in practice the full upwind scheme is frequently applied, in particular, in connection with an additional correction procedure such as Patankar's power law scheme ([15], [16]).

Now, for $w_h \in V_h := S_{hl}^d$, $u_h, v_h \in S_{hl}$ we define

$$\gamma_{ij} := m_{ij}^{-1} \sum_{T \in \mathcal{T}_h \,:\, m_{ij}^T > 0} \int\limits_{\Gamma_{ij}^T} \nu \cdot w_h \, ds, \quad r_{ij} := r\left(\frac{\gamma_{ij} d_{ij}}{\varepsilon}\right)$$

and set

$$n_{sh}(w_h, u_h, v_h) := \frac{1}{2} \sum_{i \in \Lambda} \sum_{j \in \Lambda_i} \left[\left(r_{ij} - \frac{1}{2}\right)(u_{hi} - u_{hj})(v_{hi} - v_{hj}) + \frac{1}{2}(u_{hj} v_{hi} - u_{hi} v_{hj})\right] \gamma_{ij} m_{ij}. \quad (3.1)$$

Returning to the original form n, we set for $w_h, u_h, v_h \in V_h$

$$n(w, u, v) \approx n_h(w_h, u_h, v_h) := \sum_{l=1}^{d} n_{sh}(w_h, u_h^l, v_h^l).$$

Collorary 3.1 (Semidefiniteness of n_h). *If the control function r satisfies* $(P5)$, *then there holds*

$$\forall w_h, v_h \in V_h : \quad n_h(w_h, v_h, v_h) \geqslant 0.$$

Proof. Follows immediately from the above definition (3.1) of n_{sh} and from $(P5)$. □

Finally, some discrete norms and operators have to be introduced. For $u, v \in V + V_h$, we set

$$|v|_h := \sqrt{(\nabla v, \nabla v)_h}, \qquad \|v\|_h := \sqrt{\|v\|_{0,2,\Omega}^2 + |v|_h^2}.$$

By $I_h : \overset{\circ}{W}{}_2^1(\Omega) \to S_{hl}$, the interpolation operator is denoted, whereas $L_h : C(\overline{\Omega}) + S_{hl} \to L_\infty(\Omega)$ stands for the so-called *lumping* procedure. That is, the image of L_h is a subspace consisting of functions being constant on the elements of the secondary partition $\mathcal{T}_h^*$. The application of I_h or L_h to a vector field from V or $[C(\overline{\Omega}) + S_{hl}]^d$, respectively, should be understood in a componentwise manner.

Concrete properties of these operators will be included in the subsequent assumptions. We also recall for completeness some results from [2] related to the remaining two properties of n_h, i.e., the Lipschitz-continuity and consistency.

One group of assumptions ((A13) – (A15)) establishes relations between different seminorms or norms on V_h, the other group ((A16) – (A19)) includes requirements for the operators $I_h : \overset{\circ}{W}{}_2^1(\Omega) \to S_{hl}$ and $L_h : C(\overline{\Omega}) + S_{hl} \to L_\infty(\Omega)$ mentioned at the end of Section 3.

(A13) There exists a constant $C > 0$ independent of h such that for all $v_h \in S_{hl}$

$$\sum_{i\in\Lambda}\sum_{j\in\Lambda_i}(v_{hi} - v_{hj})^2\frac{m_{ij}}{d_{ij}} \leqslant C|v_h|_h^2.$$

holds.

Collorary 3.2. *Under Assumption* (*A*13)

$$\sum_{i\in\Lambda}\sum_{j\in\Lambda_i}|w_{hj} - w_{hi}|\,|v_{hi} - v_{hj}|\frac{m_{ij}}{d_{ij}} \leqslant C|w_h|_h|v_h|_h.$$

holds.

Proof. Elementary. □

(A14) There exists a constant $C > 0$ such that for all $v_h \in S_{hl}$,

$$\|v_h\|_{0,6,\Omega} \leqslant C\|v_h\|_h.$$

holds.

Collorary 3.3. *Under Assumption* (*A*14)*, for arbitrary* $p \in [1, 6]$,

$$\|v_h\|_{0,p,\Omega} \leqslant C\|v_h\|_h.$$

holds.

Proof. Elementary. □

(A15) There exists a constant $C > 0$ such that for arbitrary $p \in [1, 6]$ and for all $v_h \in S_{hl}$,

$$\|L_h v_h\|_{0,p,\Omega} \leqslant C\|v_h\|_{0,p,\Omega}.$$

holds.

Remark 3.1. Since $\overline{\Lambda}^{\Delta}$ is nonempty, in general, $\|L_h v_h\|_{0,p,\Omega}$ is only a seminorm on S_{hl}. Both assumptions (A13) and (A15) are weakened versions of properties that do hold in many standard cases (cf. [11]). A more essential assumption is the discrete Sobolev inequality (A14). Frequently, its proof requires more involved arguments.

The above set of assumptions provides sufficient conditions for the proof of the Lipschitz-continuity of n_h (see Lemma 3.1 below). However, the rest of the assumptions about I_h and L_h will be given here, too.

(A16) There exists a constant $C > 0$ such that

(*i*) for arbitrary $p > d$ and for all $v \in W_p^1(\Omega)$, $\|(I - L_h)v\|_{0,p,\Omega} \leqslant C\,h|v|_{1,p,\Omega}$ holds,

(*ii*) for all $v_h \in S_{hl}$, $\|(I - L_h)v_h\|_{0,2,\Omega} \leqslant C\,h|v|_h$ holds.

(A17) There exists a constant $C > 0$ such that

(*i*) for all $v \in W_2^2(\Omega)$, $\|I_h v\|_{0,\infty,\Omega} \leqslant C\,\|v\|_{2,2,\Omega}$ holds,

(*ii*) for all $v \in W_2^2(\Omega)$ and all $T \in \mathfrak{T}_h$, $|I_h v|_{1,2,T} \leqslant C\,\|v\|_{2,2,T}$ holds,

(*iii*) for arbitrary $p \in (d, 6]$, $v \in W_p^1(\Omega)$ and all $T \in \mathfrak{T}_h$, $|I_h v|_{1,p,T} \leqslant C\,\|v\|_{1,p,T}$ holds.

(A18) There exists a constant $C > 0$ such that

(i) for arbitrary $p \in (d, 6]$, $v \in W^1_p(\Omega)$ and all $T \in \mathfrak{T}_h$, $|(I - I_h)v|_{l,p,T} \leqslant C\, h_T^{1-l} \|v\|_{1,p,T}$, $l = 0$ or $l = 1$, holds,

(ii) for all $v \in W^2_2(\Omega)$ and all $T \in \mathfrak{T}_h$, $|(I - I_h)v|_{l,2,T} \leqslant C\, h_T^{2-l} \|v\|_{2,2,T}$, $l = 0$ or $l = 1$, holds.

(A19) There exists a constant $C > 0$ such that for all $v \in W^2_2(\Omega)$ and all $T \in \mathfrak{T}_h$,

$$\|L_h(I - I_h)v\|_{0,2,T} \leqslant C\, h_T^2 \|v\|_{2,2,T}.$$

holds.

All the assumptions (A16) – (A19) are rather typical properties of the interpolation and lumping operators I_h and L_h, respectively. They do not imply essential restrictions on the applicability of the method. In particular, in many standard finite volume methods, the left-hand side of inequality (A19) simply vanishes.

Lemma 3.1. *Suppose* (*A*8) – (*A*15). *Then, for arbitrary* $w_h, z_h \in V_h$ *and* $u_h, v_h \in S_{hl}$ *the estimate*

$$|n_{sh}(w_h, u_h, v_h) - n_{sh}(z_h, u_h, v_h)| \leqslant C \|w_h - z_h\|_h \|u_h\|_h \|v_h\|_h$$

holds, where $C > 0$ *is a constant which does not depend on* h.

Proof. See [2, Lemma 3]. □

Lemma 3.2. *Suppose* (*A*8), (*A*11), (*A*13) – (*A*15), (*A*16) (*ii*), (*A*(17) – (*A*19). *Then, for any* $w \in W^2_2(\Omega)^d \bigcap V$ *satisfying* $\nabla \cdot w = 0$, *any* $u \in W^2_2(\Omega) \bigcap \overset{\circ}{W}{}^1_2(\Omega)$ *and any element* $v_h \in S_{hl}$, *the estimate*

$$|n_s(w, u, v_h) - n_{sh}(I_h w, I_h u, v_h)| \leqslant Ch \|w\|_{2,2,\Omega} \|u\|_{2,2,\Omega} \|v_h\|_h$$

holds, where $C > 0$ *is a constant which does not depend on* h.

Proof. See [2, Lemma 4]. □

4. Discretization by quadrilateral/hexahedral elements

4.1. Definition of the finite element space. We consider one of the four finite elements introduced by Schieweck [19], namely the so-called P_1*-parametric element,* and apply the above discretization method to the trilinear form n. The resulting discrete form n_h differs from Schieweck's one in two aspects: The internal geometry of the control volumes Ω_i and the choice of upwind parameters are different.

Each partition $\mathfrak{T}_h$, $h \in (0, h_0]$, $h_0 > 0$, of $\Omega \subset \mathbf{R}^d$ consists of convex quadrilaterals ($d = 2$) or hexahedra ($d = 3$), where the faces (i.e., the two-dimensional boundary surfaces) of the hexahedra are plane.

We define

$$\rho_T := \sup \bigl\{ \rho \,:\, B(x_T, \rho) \subset T \bigr\}.$$

Remark 4.1. In his original work [19], Schieweck did not require the convexity of $T \in \mathfrak{T}_h$. He defined ρ_T by

$$\rho_T := \sup \bigl\{ \rho \,:\, T \text{ is star-shaped w.r.t. } B(x_T, \rho) \bigr\}.$$

By $\underline{h}_T$, the length of the shortest edge of T is denoted. Finally, let $\underline{\alpha}_T$ be the smallest angle between neighbouring edges and, for $d = 3$, also between neighbouring faces, and let $\overline{\alpha}_T$ be the corresponding largest angle.

Definition 4.1. The family $\mathcal{F} := \{\mathcal{T}_h\}_{h\in(0,h_0]}$ is called *shape-regular* iff there exist positive constants $\gamma_0, \gamma_1, \alpha_0$ independent of any particular element $T \in \mathcal{T}_h$ such that

$$\forall \mathcal{T}_h \in \mathcal{F} \quad \forall T \in \mathcal{T}_h : \quad \frac{h_T}{\underline{h}_T} \leqslant \gamma_0, \quad \frac{h_T}{\rho_T} \leqslant \gamma_1, \quad 0 < \alpha_0 \leqslant \underline{\alpha}_T \leqslant \overline{\alpha}_T \leqslant \pi - \alpha_0.$$

holds.

Now, let $T \in \mathcal{T}_h$ be fixed. Denote by $\mathcal{E}_T$ the set of all its faces and let $x_E \in E$ be the barycentre of the face $E \in \mathcal{E}_T$. Using local enumeration of the $2d$ faces $E \in \mathcal{E}_T$ such that E_j and E_{j+d} are opposite to each other, $j \in [1, d]_{\mathbf{N}}$, it is possible to define an affine mapping F_T in such a way that

$$F_T(\hat{e}_j) = x_{E_j}, \quad F_T(-\hat{e}_j) = x_{E_{j+d}},$$

holds, where $\hat{e}_j$ denotes the j-th canonical unit vector in $\mathbf{R}^d$.

It is not difficult to give this mapping explicitely. In fact, if $B_T \in \mathbf{R}^{d,d}$ is a matrix with column vectors $b_{Tj} := x_{E_j} - x_T$, then the transformation $F_T(\hat{x}) := B_T\hat{x} + x_T$ possesses the indicated properties.

The fundamental difficulty in the analysis of this approach is, unfortunately, that the "reference" element $\hat{T}_T := F_T^{-1}(T)$ does not form the unit d-cube, in general. The only statement that can be given about the "family of reference elements" is as follows.

Remark 4.2. All the reference elements $\hat{T}_T$ have the same face barycenters (and, therefore, the same element barycentre $\hat{0}$).

Proof. Simple calculation. □

Lemma 4.1. *Assume that $\mathcal{F}$ is shape-regular. Then there exist constants depending only on $\mathcal{F}$ such that*

$$\forall \mathcal{T}_h \in \mathcal{F} \quad \forall T \in \mathcal{T}_h : \quad \|B_T\| \leqslant Ch_T, \quad \|B_T^{-1}\| \leqslant Ch_T^{-1}, \quad c^{-1}h_T^d \leqslant \det B_T \leqslant c^{-1}h_T^d,$$

holds, where $\|\cdot\|$ is an arbitrary but fixed matrix norm.

Proof. Omitted. □

Collorary 4.1. *Under the assumptions of the above Lemma* 4.1*, there exists some number $\rho > 0$ such that*

$$\forall \mathcal{T}_h \in \mathcal{F} \; \forall T \in \mathcal{T}_h : \; \hat{T}_T = F_T^{-1}(T) \subset \hat{B}(0, \rho).$$

Proof. By the lemma,

$$\|\hat{x}\| = \|B_T^{-1}(x - x_T)\| \leqslant \|B_T^{-1}\|h_T \leqslant C =: \rho.$$

Now we turn to the description of the finite element space. Starting with the local basis on $\hat{T}_T$, it will be transformed to the original element T by means of F_T.

We set

$$\hat{\mathcal{P}} := \operatorname*{span}_{j\in[1,d]_{\mathbf{N}},\, k\in[1,d-1]_{\mathbf{N}}} \left\{1, \hat{x}_j, \hat{x}_k^2 - \hat{x}_{k+1}^2\right\} \tag{4.1}$$

Obviously, $\dim(\hat{\mathcal{P}}) = 2d$. For simplicity in what follows, the basis polynomials used in (4.1) are denoted by $\hat{p}_k$, $k \in [1, 2d]_{\mathbf{N}}$ (as they appear there). To define the local basis elements (shape functions), the following functionals on $\hat{\mathcal{P}}$ (local degrees of freedom) are used:

$$\hat{\Phi}_j(\hat{\varphi}) := \hat{\varphi}(\hat{x}_{\hat{E}_j}),$$

where $\hat{x}_{\hat{E}_j} := \hat{e}_j$, $\hat{x}_{\hat{E}_{j+d}} := -\hat{e}_j$, $j \in [1,d]_{\mathbf{N}}$. Then the system $\hat{\Phi}_j(\hat{\varphi}_i) = \delta_{ij}$, $i,j \in [1,2d]_{\mathbf{N}}$, gives $2d$ conditions for the canonical determination of $\{\hat{\varphi}_i\}_{i\in[1,2d]_{\mathbf{N}}}$. If the representation $\hat{\varphi}_i(\hat{x}) = \sum_{k\in[1,2d]_{\mathbf{N}}} c_k^{(i)} \hat{p}_k(\hat{x})$ is used, then due to the above Remark 4.2 the matrix

$$(\hat{p}_k(\hat{x}_{\hat{E}_j}))_{j,k\in[1,2d]_{\mathbf{N}}} \tag{4.2}$$

of the system is the same for all elements $\hat{T}_T$. Therefore, in order to prove that the pairing $(\{\hat{\varphi}_i\}, \{\hat{\Phi}_j\})$ is $\hat{\mathcal{P}}$-unisolvent, it is sufficient to consider a particular situation, namely the unit cube $\hat{T} := (-1,1)^d$. Then it easily turns out that the matrix (4.2) of the system is regular.

Furthermore, we set $Q_h := \{q_h \in Q : \ q_h|_T \in \mathcal{P}_0 \ (\forall T \in \mathcal{T}_h)\}$.

4.2. Discretization of the trilinear form. For any $T \in \mathcal{T}_h$ and any $i \in \Lambda_T$, we have to define the contributions Ω_i^T to the element $\Omega_i \in \mathcal{T}_h^*$ associated with the local nodes $x_i \in \partial T$.

To meet Assumption (A8), we first of all describe the subdomains $\Omega_{ij}^{T,l}$.

So let x_{E_i}, x_{E_j} be the barycenters of two neighbouring faces E_i, E_j (i.e., they have in common $d-1$ vertices $y_{ij}^{T,k}$ of T, $k \in [1,d-1]_{\mathbf{N}}$, and consider the midpoint x_{ij} of the line segment connecting both points, i.e., $x_{ij} := (x_i + x_j)/2$. Since T is convex, $x_{ij} \in \overline{T}$. Then x_{ij}, x_T and the common vertices determine d $(d-1)$-simplices which have the point x_{ij} in common and form the boundary parts $\Gamma_{ij}^{T,l}$.

Namely, for $d = 2$ (see Fig. 4.1, left),

$$\Gamma_{ij}^{T,1} := \operatorname{conv}\{x_{ij}, y_{ij}^{T,1}\}, \quad \Gamma_{ij}^{T,2} := \operatorname{conv}\{x_{ij}, x_T\};$$

for $d = 3$ (see Fig. 4.2),

$$\Gamma_{ij}^{T,1} := \operatorname{conv}\{x_{ij}, x_T, y_{ij}^{T,1}\}, \quad \Gamma_{ij}^{T,2} := \operatorname{conv}\{x_{ij}, x_T, y_{ij}^{T,2}\},$$

$$\Gamma_{ij}^{T,3} := \operatorname{conv}\{x_{ij}, y_{ij}^{T,1}, y_{ij}^{T,2}\}.$$

The convex hull of x_i, x_j with each of these boundary parts generate the subdomains $\Omega_{ij}^{T,l}$.

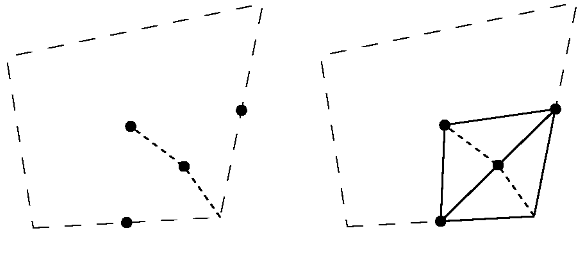

Fig. 4.1. The case of $d = 2$: $\Gamma_{ij}^{T,1}$, $\Gamma_{ij}^{T,2}$ (left, finely dotted lines) and $\Omega_{ij}^{T,1}$, $\Omega_{ij}^{T,2}$ (right, bounded by continuous lines)

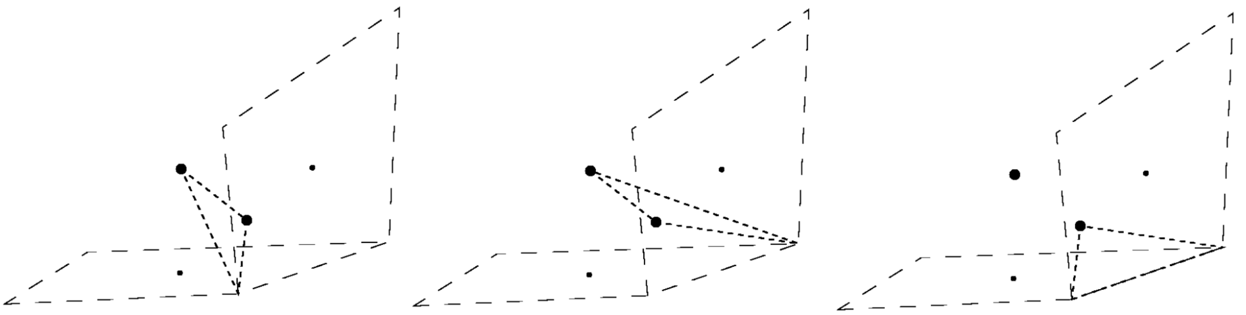

Fig. 4.2. Boundary parts $\Gamma_{ij}^{T,l}$ in the case of $d = 3$ (finely dotted lines)

That is, for $d = 2$,

$$\Omega_{ij}^{T,1} := \operatorname{int}(\operatorname{conv}\{x_i, x_j, y_{ij}^{T,1}\}), \quad \Omega_{ij}^{T,2} := \operatorname{int}(\operatorname{conv}\{x_i, x_j, x_T\});$$

for $d = 3$ (see Fig. 4.3),

$$\Omega_{ij}^{T,1} := \operatorname{int}(\operatorname{conv}\{x_i, x_j, x_T, y_{ij}^{T,1}\}), \quad \Omega_{ij}^{T,2} := \operatorname{int}(\operatorname{conv}\{x_i, x_j, x_T, y_{ij}^{T,2}\}),$$

$$\Omega_{ij}^{T,3} := \operatorname{int}(\operatorname{conv}\{x_i, x_j, y_{ij}^{T,1}, y_{ij}^{T,2}\}).$$

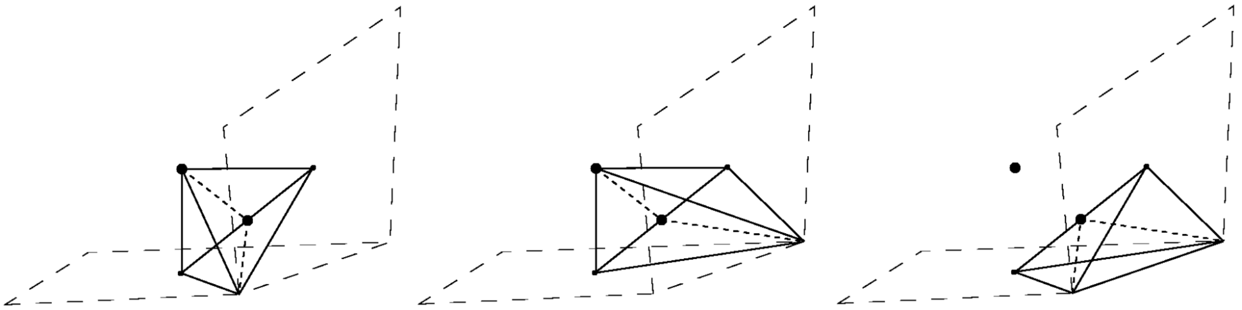

Fig. 4.3. Subdomains $\Omega_{ij}^{T,l}$ in the case of $d = 3$ (bounded by continuous lines)

4.3. Verification of the assumptions. First of all we will show that the local quasi-uniformity of the family of triangulations generated by the simplices $\Omega_{ij}^{T,l}$ (see Assumption (A8)) is a sufficient condition to satisfy the collection of geometrical assumptions (A9)–(A12).

That is, we assume that there exists a constant $c > 0$ depending only on $\mathcal{F}$ such that

$$\forall\, \Omega_{ij}^{T,l}: \quad c^{-1}(h_{ij}^{T,l})^d \leqslant \operatorname{meas}_d(\Omega_{ij}^{T,l}) \leqslant c(h_{ij}^{T,l})^d \tag{4.3}$$

holds.

4.3.1. Assumption (A9). We show that $d_{ij}^5 \leqslant C m_{ij}^{T,l}$; then the statement immediately follows from the trivial estimate[1] $m_{ij}^{T,l} \leqslant m_{ij}$.

From (4.3) we conclude that the length of any edge of the simplex $\Omega_{ij}^{T,l}$ lies within the interval[2] $[\tilde{c}h_{ij}^{T,l}, h_{ij}^{T,l}]$, where $\tilde{c}$ depends on c from (4.3). In particular,

$$d_{ij}^{T,l} \in [\tilde{c}h_{ij}^{T,l}, h_{ij}^{T,l}]. \tag{4.4}$$

In the case of $d = 2$, $m_{ij}^{T,l}$ is not smaller than the height of the triangle corresponding to the edge connecting x_i to x_j, and that height is not smaller than the diameter of the inscribed circle, i.e.,

$$\tilde{c}h_{ij}^{T,l} \leqslant m_{ij}^{T,l}. \tag{4.5}$$

So we get $d_{ij} \leqslant h_{ij}^{T,l} \leqslant \tilde{c}^{-1} m_{ij}^{T,l}$. Since we may assume, without loss of generality, that $d_{ij} \leqslant 1$, the statement follows.

In the case of $d = 3$, we introduce the temporal notation $y_{ij}^{T,d} := x_T$ and observe that the central projection of the inscribed ball from the vertex to the opposite face gives a circle

[1] In fact, it is sufficient that in the set $\{(T,l)\}_{T\in\mathcal{T}_h: m_{ij}^T > 0,\, l\in[1, l_{ij}^{T,l}]_{\mathbf{N}}}$ there exists one element for which the relation $d_{ij}^5 \leqslant C m_{ij}^{T,l}$ is valid.

[2] $\tilde{c}h_{ij}^{T,l}$ is the diameter of the inscribed ball.

which is contained in that face. Therefore, the length of the edges of the face as well as the face-heights corresponding to these edges are not shorter than the diameter of the circle which itself is not smaller than the diameter of the inscribed ball.

The area $m_{ij}^{T,l}$ of the triangle $\Gamma_{ij}^{T,l}$ is equal to the half of the length of the line segment connecting x_{ij} with one of the vertices $y_{ij}^{T,l}$ multiplied by the height in the triangle $\Gamma_{ij}^{T,l}$.

The first factor is not smaller than the height of the face formed by x_i, x_j and $y_{ij}^{T,l}$ corresponding to the base line given by x_i, x_j; hence it is not smaller than $\tilde{c}h_{ij}^{T,l}$.

The second factor is not smaller than the height of the tetrahedron $\Omega_{ij}^{T,l}$ corresponding to the face, i.e. it is not smaller than the diameter of the inscribed ball either.

In the final analysis, we get

$$\frac{1}{2}(\tilde{c}h_{ij}^{T,l})^2 \leqslant m_{ij}^{T,l}, \tag{4.6}$$

thus $d_{ij}^2 \leqslant (h_{ij}^{T,l})^2 \leqslant 2\tilde{c}^{-2}m_{ij}^{T,l}$. From this relation the statement follows.

4.3.2. Assumption (A10). We show that $d_{ij}(h_{ij}^{T,l})^{d-1} \leqslant Cm_{ij}^{T,l}$. But this is a consequence of the above relations (4.4), (4.5) and (4.6). Indeed, for $d = 2$ the first two estimates imply $d_{ij}h_{ij}^{T,l} \leqslant h_{ij}^{T,l} \leqslant \tilde{c}^{-1}m_{ij}^{T,l}$, whereas for $d = 3$, (4.4) and (4.6) they yield

$$d_{ij}(h_{ij}^{T,l})^2 \leqslant (h_{ij}^{T,l})^2 \leqslant 2\tilde{c}^{-2}m_{ij}^{T,l}.$$

4.3.3. Assumption (A11). Obviously, we have

$$m_{ij}^{T,l} \leqslant (h_{ij}^{T,l})^{d-1}. \tag{4.7}$$

It follows by (4.4) $m_{ij}^{T,l}d_{ij} \leqslant (h_{ij}^{T,l})^d$, and (4.3) results in $m_{ij}^{T,l}d_{ij} \leqslant \operatorname{meas}_d(\Omega_{ij}^{T,l})$.

4.3.4. Assumption (A12). This relation is a consequence of (4.3):

$$(h_{ij}^{T,l})^4 \leqslant (h_{ij}^{T,l})^d \leqslant C\operatorname{meas}_d(\Omega_{ij}^{T,l}).$$

In the next step, the embedding relations (A13) and (A14) will be verified.

4.3.5. Assumption (A13). This assumption can also be verified by using (4.3). Namely, we decompose the double sum that appears on the left-hand side in a finer manner as

$$\sum_{i\in\Lambda}\sum_{j\in\Lambda_i}(v_{hi}-v_{hj})^2\frac{m_{ij}}{d_{ij}} = \sum_{i\in\Lambda}\sum_{j\in\Lambda_i}\sum_{T\in\mathcal{T}_h\,:\,m_{ij}^T>0}\ \sum_{l\in[1,l_{ij}^T]_N}(v_{hi}-v_{hj})^2\frac{m_{ij}^{T,l}}{d_{ij}}.$$

Now we transform $\Omega_{ij}^{T,l}$ into the reference simplex and get (for the restrictions on v_h into $\Omega_{ij}^{T,l}$)

$$|v_{hi}-v_{hj}| = |\hat{v}_{hi}-\hat{v}_{hj}| = \left|(\hat{x}_i-\hat{x}_j)\int_0^1 \hat{\nabla}\hat{v}_h(\hat{x}_j+s(\hat{x}_i-\hat{x}_j))\,ds\right| \leqslant \|\hat{x}_i-\hat{x}_j\|\,\|\hat{\nabla}\hat{v}_h\|_{0,\infty,\hat{T}}.$$

Since $\hat{\nabla}\hat{v}_h \in [\mathcal{P}_1(\hat{T})]^d$, the norms $\|\hat{\nabla}\hat{v}_h\|_{0,\infty,\hat{T}}$ and $\|\hat{\nabla}\hat{v}_h\|_{0,2,\hat{T}}$ are equivalent (as norms on the finite-dimensional space $[\mathcal{P}_1(\hat{T})]^d$). Consequently, $|v_{hi}-v_{hj}| \leqslant C\|\hat{\nabla}\hat{v}_h\|_{0,2,\hat{T}}$, and the back-transformation implies

$$|v_{hi}-v_{hj}| \leqslant C\frac{h_{ij}^{T,l}}{\operatorname{meas}_d(\Omega_{ij}^{T,l})^{1/2}}\|\nabla v_h\|_{0,2,\Omega_{ij}^{T,l}}.$$

Since it holds, by (4.4), (4.7), (4.3),

$$\frac{m_{ij}^{T,l}(h_{ij}^{T,l})^2}{d_{ij}\mathrm{meas}_d(\Omega_{ij}^{T,l})} \leqslant \frac{1}{\tilde{c}}\frac{m_{ij}^{T,l}h_{ij}^{T,l}}{\mathrm{meas}_d(\Omega_{ij}^{T,l})} \leqslant \frac{1}{\tilde{c}}\frac{(h_{ij}^{T,l})^d}{d_{ij}\mathrm{meas}_d(\Omega_{ij}^{T,l})} \leqslant \frac{c}{\tilde{c}},$$

we conclude that

$$\sum_{i\in\Lambda}\sum_{j\in\Lambda_i}(v_{hi}-v_{hj})^2\frac{m_{ij}}{d_{ij}} \leqslant C\sum_{i\in\Lambda}\sum_{j\in\Lambda_i}\sum_{T\in\mathcal{T}_h\,:\,m_{ij}^T>0}\|\nabla v_h\|^2_{0,2,\Omega_{ij}^T} \leqslant C\sum_{T\in\mathcal{T}_h}\|\nabla v_h\|^2_{0,2,T} = C|v_h|_h^2.$$

4.3.6. Assumption (A14). Verification of this assumption completely follows the lines of [19, Lemma 4.4 there]. That is, based on the design of the auxiliary interpolation operator acting in a *conforming* finite element space, the sharper estimate

$$\|v_h\|_{0,p,\Omega} \leqslant C|v_h|_h \tag{4.8}$$

is obtained by using the standard embedding result (w.r.t. the conforming part) and a suitable interpolation error estimate. The result is valid under the assumption that the basic family $\mathcal{F}$ of partitions is shape-regular (cf. Definition 4.1).

4.3.7. Assumption (A15). We introduce the following lumping operator $L_h : C(\overline{\Omega}) + S_{hl} \to L_\infty(\Omega)$ via

$$v \mapsto L_h v := \sum_{i\in\overline{\Lambda}} v(x_i)\chi_{\Omega_i}, \tag{4.9}$$

where χ_{Ω_i} is the indicator function of the element Ω_i of the secondary partition $\mathcal{T}_h^*$. This defines

$$(\cdot,\cdot)_l := (L_h\cdot, L_h\cdot). \tag{4.10}$$

Then we can write

$$\|L_h v_h\|^p_{0,p,\Omega} = \sum_{i\in\Lambda}\sum_{j\in\Lambda_i}\sum_{T\in\mathcal{T}_h\,:\,m_{ij}^T>0}\sum_{l\in[1,l_{ij}^T]_N}\int\limits_{\Omega_{ij}^{T,l}}|L_h v_h|^p\,dx.$$

Now the following estimate is trivial: $\int_{\Omega_{ij}^{T,l}}|L_h v_h|^p\,dx \leqslant \int_{\Omega_{ij}^{T,l}}|v_{hi}|^p\,dx + \int_{\Omega_{ij}^{T,l}}|v_{hj}|^p\,dx$. Transforming the integrals into the reference element $\hat{T}$, we get for the first integral

$$\int\limits_{\Omega_{ij}^{T,l}}|v_{hi}|^p dx = \frac{\mathrm{meas}_d(\Omega_{ij}^{T,l})}{\mathrm{meas}_d(\hat{T})}\int\limits_{\hat{T}}|\hat{v}_{hi}|^p d\hat{x}.$$

On $\hat{\mathcal{P}}$, the mapping $\hat{\varphi} \mapsto \{\int_{\hat{T}}|\hat{\varphi}(\hat{x}_i)|^p\,d\hat{x}\}^{1/p}$ is a seminorm (cf. Assumption (A8) (*ii*)). Since in a finite-dimensional space a seminorm can be estimated by the norm (up to some multiplicative constant), there exists a constant $C > 0$ such that

$$\int\limits_{\hat{T}}|\hat{v}_{hi}|^p d\hat{x} \leqslant C\frac{\mathrm{meas}_d(\Omega_{ij}^{T,l})}{\mathrm{meas}_d(\hat{T})}\int\limits_{\hat{T}}|\hat{v}_h(\hat{x})|^p d\hat{x} \leqslant C\|v_h\|^p_{0,p,\Omega_{ij}^{T,l}}.$$

A similar estimate is true for the second integral. After summation, we arrive at the desired relation.

4.3.8. Assumption (A16).

$$\|(I-L_h)v\|_{0,p,\Omega}^p = \sum_{i\in\Lambda}\sum_{j\in\Lambda_i}\sum_{T\in\mathfrak{T}_h\,:\,m_{ij}^T>0}\sum_{l\in[1,l_{ij}^T]_N}\int\limits_{\Omega_{ij}^{T,l}}|(I-L_h)v|^p dx$$

holds. Now we have

$$\int\limits_{\Omega_{ij}^{T,l}}|(I-L_h)v|^p dx = \int\limits_{\Omega_{ij}^{T,l}\bigcap\Omega_i}|v-v_i|^p dx + \int\limits_{\Omega_{ij}^{T,l}\bigcap\Omega_j}|v-v_j|^p dx \leqslant \int\limits_{\Omega_{ij}^{T,l}}|v-v_i|^p dx + \int\limits_{\Omega_{ij}^{T,l}}|v-v_j|^p dx.$$

Due to symmetry relation $\Omega_{ij}^{T,l}=\Omega_{ji}^{T,l}$, see Assumption (A8) (i), it is sufficient to estimate one of the integrals.

By Assumption (A8), we can write, for any fixed $w\in L_q(\Omega_{ij}^{T,l})$, $q=p/(p-1)$, that

$$\int\limits_{\Omega_{ij}^{T,l}}(v-v_i)w\,dx = \frac{\mathrm{meas}_d(\Omega_{ij}^{T,l})}{\mathrm{meas}_d(\hat{T})}\int\limits_{\hat{T}}(\hat{v}-\hat{v}_i)\hat{w}\,d\hat{x}.$$

It follows that

$$\left|\int\limits_{\hat{T}}(\hat{v}-\hat{v}_i)\hat{w}\,d\hat{x}\right| \leqslant \left[\|\hat{v}\|_{0,p,\hat{T}}+\|\hat{v}_i\|_{0,p,\hat{T}}\right]\|\hat{w}\|_{0,q,\hat{T}}.$$

Since $\|\hat{v}_i\|_{0,p,\hat{T}}\leqslant\sqrt{\mathrm{meas}_d(\hat{T})}\|\hat{v}\|_{0,\infty,\hat{T}}$ and, by Sobolev's embedding theorem, $\|\hat{v}\|_{0,\infty,\hat{T}}\leqslant C\|\hat{v}\|_{1,p,\hat{T}}$, we obtain

$$\left|\int\limits_{\hat{T}}(\hat{v}-\hat{v}_i)\hat{w}\,d\hat{x}\right| \leqslant C\|\hat{v}\|_{1,p,\hat{T}}\|\hat{w}\|_{0,q,\hat{T}}.$$

The integral on the left-hand side is a linear continuous functional of the argument $\hat{v}\in W_p^1(\hat{T})$, and it vanishes for constant arguments. Hence the Bramble — Hilbert lemma implies that

$$\left|\int\limits_{\hat{T}}(\hat{v}-\hat{v}_i)\hat{w}\,d\hat{x}\right| \leqslant C|\hat{v}|_{1,p,\hat{T}}\|\hat{w}\|_{0,q,\hat{T}},$$

and the back-transformation results in the estimate

$$\left|\int\limits_{\Omega_{ij}^{T,l}}(v-v_i)w\,dx\right| \leqslant Ch_{ij}^{T,l}|v|_{1,p,\Omega_{ij}^{T,l}}\|w\|_{0,q,\Omega_{ij}^{T,l}}.$$

Therefore,

$$\|v-v_i\|_{1,p,\Omega_{ij}^{T,l}} = \sup_{w\in L_q(\Omega_{ij}^{T,l})}\frac{(v-v_i,w)_{\Omega_{ij}^{T,l}}}{\|w\|_{0,q,\Omega_{ij}^{T,l}}} \leqslant Ch_{ij}^{T,l}|v|_{1,p,\Omega_{ij}^{T,l}}.$$

In the final analysis, we get the first estimate

$$\|(I-L_h)v\|_{0,p,\Omega}^p \leqslant Ch^p\sum_{T\in\mathfrak{T}_h}|v|_{1,p,T}^p = Ch^p|v|_{1,p,\Omega}^p.$$

To verify (ii), we proceed as in the first part and arrive at the estimate

$$\left| \int\limits_{\hat T} (\hat v_h - \hat v_{hi}) \hat w \, d\hat x \right| \leqslant \left[\|\hat v_h\|_{0,p,\hat T} + \|\hat v_{hi}\|_{0,p,\hat T} \right] \|\hat w\|_{0,q,\hat T}.$$

Since $\|\hat v_{hi}\|_{0,p,\hat T}$ is a seminorm on $\hat{\mathcal{P}}$, we obtain

$$\left| \int\limits_{\hat T} (\hat v_h - \hat v_{hi}) \hat w \, d\hat x \right| \leqslant C \|\hat v_h\|_{0,p,\hat T} \|\hat w\|_{0,q,\hat T} \leqslant C \|\hat v_h\|_{1,p,\hat T} \|\hat w\|_{0,q,\hat T}.$$

The integral on the left-hand side is a linear continuous functional of the argument $\hat v_h \in \hat{\mathcal{P}}$ and it vanishes for constant arguments. Hence the Bramble — Hilbert lemma implies that

$$\left| \int\limits_{\hat T} (\hat v_h - \hat v_{hi}) \hat w \, d\hat x \right| \leqslant C |\hat v_h|_{1,p,\hat T} \|\hat w\|_{0,q,\hat T}.$$

and the back-transformation results in the estimate

$$\left| \int\limits_{\Omega_{ij}^{T,l}} (v_h - v_{hi}) w \, dx \right| \leqslant C h_{ij}^{T,l} |v_h|_{1,p,\Omega_{ij}^{T,l}} \|w\|_{0,q,\Omega_{ij}^{T,l}}.$$

Therefore, $\|v_h - v_{hi}\|_{1,p,\Omega_{ij}^{T,l}} \leqslant C h_{ij}^{T,l} |v_h|_{1,p,\Omega_{ij}^{T,l}}$. In the final analysis, we get

$$\|(I - L_h) v_h\|_{0,p,\Omega}^p \leqslant C h^p \sum_{T \in \mathcal{T}_h} |v_h|_{1,p,T}^p,$$

which gives for $p = 2$ the desired result.

4.3.9. Assumption (A17). The interpolation operator $I_h : V \to V_h$ is defined according to [19, Definition 4.5], i.e.,

$$I_h v := \sum_{i \in \overline{\Lambda}_g} v_i \varphi_i \quad \text{with} \quad v_i := \frac{1}{\text{meas}_{d-1}(E_i)} \int\limits_{E_i} v \, ds.$$

It is clearly sufficient to consider only one component. As a consequence, we get for $v \in W_2^2(\Omega)$

$$\|I_h v\|_{0,\infty,\Omega} = \max_{T' \in \mathcal{T}_h} \|I_h v\|_{0,\infty,T'} = \|I_h v\|_{0,\infty,T}$$

for some element $T \in \mathcal{T}_h$. Now,

$$\|I_h v\|_{0,\infty,T} \leqslant \sum_{i \in \overline{\Lambda}_g} |v_i| \|\varphi_i\|_{0,\infty,T} \leqslant \left(\max_{i \in \overline{\Lambda}_g} |v_i| \right) \sum_{i \in \overline{\Lambda}_g} \|\varphi_i\|_{0,\infty,T}.$$

holds. From $|v_i| \leqslant \|v\|_{C(E_i)} \leqslant \|v\|_{C(\overline{\Omega})} \leqslant C \|v\|_{2,2,\Omega}$ we get $\|I_h v\|_{0,\infty,T} \leqslant C \|v\|_{2,2,\Omega}$, thus,

$$\|I_h v\|_{0,\infty,\Omega} \leqslant C \|v\|_{2,2,\Omega}.$$

Relation (ii) of the assumption is a simple consequence of Assumption (A18) (ii) :

$$|I_h v|_{1,2,T} \leqslant |v|_{1,2,T} + |(I - I_h)v|_{1,2,T} \leqslant [1 + C\tilde{h}_T]\|v\|_{2,2,T},$$

where

$$\tilde{h}_T := h_T + \alpha_T, \tag{4.11}$$

and α_T is defined as follows: For $d = 2$, it is the largest angle between two *opposite* edges. For $d = 3$, such a quantity, say α_E now can be defined for any face E of T. Then we set

$$\alpha_T := \max_{E \in \partial T} \alpha_E.$$

To verify (iii), we use a similar argument. In view of (A18) (i), we have

$$|I_h v|_{1,p,T} \leqslant |v|_{1,p,T} + |(I - I_h)v|_{1,p,T} \leqslant [1 + C]\|v\|_{1,p,T}.$$

4.3.10. Assumption (A18). To prove (i), we first of all introduce two additional interpolation operators. The first one is the natural interpolation operator $I_{V_h} : \left[W_p^1(\Omega)\right]^d \bigcap V \to V_h$ defined via

$$I_{V_h} v := \sum_{i \in \overline{\Lambda}_g} v(x_i)\varphi_i.$$

Again, it is sufficient to consider one component of it. Moreover, we use for any $v \in W_p^1(\Omega)$ the linear interpolant $I_1 v$ of v on T (for example, an averaged Taylor polynomial [6, Ch. 4]), for which we have the estimate [6, Lemma 4.3.8],

$$|(I - I_1)v|_{l,p,T} \leqslant C h_T^{1-l} |v|_{1,p,T}, \quad v \in W_p^1(\Omega), \quad l = 0 \ \text{ or } \ l = 1. \tag{4.12}$$

Now we make use of the following simple inequality:

$$|(I - I_h)v|_{l,p,T} \leqslant |(I - I_1)v|_{l,p,T} + |(I - I_{V_h})I_1 v|_{l,p,T} + |(I_{V_h} - I_h)I_1 v|_{l,p,T} + |I_h(I_1 - I)v|_{l,p,T}.$$

Due to (4.12), it remains to estimate the last three terms. It is not difficult to verify (by a slight modification of the proof of [19, Lemma 2.14]) that

$$|(I - I_{V_h})I_1 v|_{l,p,T} \leqslant C h_T^{1-l} |I_1 v|_{1,p,T}.$$

holds. Now, in order to keep the presentation clear, let w be the restriction on $I_1 v$ into T, i.e., $w := I_1 v\big|_T$. For $x \in T$,

$$(I_{V_h} - I_h)w(x) = \sum_{i \in \Lambda_{gT}} [w(x_i) - w_i]\varphi_i(x)$$

holds, where $w_i = \{\text{meas}_{d-1}(E_i)\}^{-1} \int_{E_i} w\, ds$. In the case of $d = 2$, the midpoint rule integrates the linear polynomials exactly, hence $w(x_i) - w_i = 0$ simply holds. In the case of $d = 3$, for any $w \in \mathcal{P}_1(T)$, we have

$$w(x) = w(x_i) + \nabla w(x_i) \cdot (x - x_i)$$

on T since ∇w is constant on T.

It follows that

$$|w(x_i) - w_i| = \left| \frac{1}{\text{meas}_{d-1}(E_i)} \int\limits_{E_i} \nabla w(x_i) \cdot (x - x_i) ds \right| \leqslant h_T \|\nabla w(x_i)\| \leqslant$$

$$h_T|w|_{1,\infty,T} \leqslant h_T\{\mathrm{meas}_d(T)\}^{-1/p}|w|_{1,p,T}.$$

Thus, we arrive at

$$|(I_{V_h} - I_h)w|_{l,p,T} \leqslant Ch_T\left\{\mathrm{meas}_d(T)\right\}^{-1/p}|w|_{1,p,T}\sum_{i\in\Lambda_{gT}}|\varphi_i|_{l,p,T}.$$

The estimate

$$|\varphi_i|_{l,p,T} \leqslant Ch_T^{-l}\left\{\mathrm{meas}_d(T)\right\}^{1/p}|\varphi_i|_{0,\infty,T} \leqslant Ch_T^{-l}\left\{\mathrm{meas}_d(T)\right\}^{1/p}$$

implies

$$|(I_{V_h} - I_h)w|_{l,p,T} \leqslant Ch_T^{1-l}|w|_{1,p,T}.$$

It remains to take into consideration the above estimate (4.12) for I_1 to get

$$|w|_{1,p,T} \leqslant |v|_{1,p,T} + |(I - I_1)v|_{1,p,T} \leqslant C|v|_{1,p,T},$$

hence

$$|(I_{V_h} - I_h)w|_{l,p,T} \leqslant Ch_T^{1-l}|v|_{1,p,T}.$$

In order to prove the fourth estimate, we first consider the stability of I_h in $W_p^l(\Omega)$. By a slight modification of the proof of [2, Lemma 2], where it is necessary to take into consideration the fact (see [7]) that we have no unique reference element, but the constant in the Bramble-Hilbert lemma only depends on the diameter of $\hat{T}_T$ which can be bounded for all $\hat{T}_T$, we have for $w \in W_p^1(T)$

$$\left|\frac{1}{\mathrm{meas}_{d-1}(E_i)}\int\limits_{E_i} w\,ds\right| \leqslant \frac{1}{\mathrm{meas}_d(T)}\left[\|w\|_{0,1,T} + Ch_T|w|_{0,1,T}\right] \leqslant$$

$$\mathrm{meas}_d(T)^{-1/p}\left[\|w\|_{0,p,T} + Ch_T|w|_{0,p,T}\right].$$

Consequently (cf. a similar argument above),

$$|I_hw|_{l,p,T} \leqslant \mathrm{meas}_d(T)^{-1/p}\left[\|w\|_{0,p,T} + Ch_T|w|_{0,p,T}\right]\sum_{i\in\Lambda_{gT}}|\varphi_i|_{l,p,T} \leqslant$$

$$\left[\|w\|_{0,p,T} + Ch_T|w|_{0,p,T}\right]\sum_{i\in\Lambda_{gT}}|\varphi_i|_{l,\infty,T}.$$

With the particular choice $w := (I - I_1)v\big|_T$ and the above error estimate (4.12) we obtain

$$|I_h(I_1 - I)v|_{l,p,T} \leqslant Ch_T|v|_{l,p,T}\sum_{i\in\Lambda_{gT}}|\varphi_i|_{l,\infty,T} \leqslant Ch_T^{1-l}|v|_{l,p,T}.$$

Thus, the desired estimate has been proved.

To verify (ii), we simply refer to [19, Lemma 4.10]:

$$|(I - I_h)v|_{l,2,T} \leqslant C\tilde{h}_Th_T^{1-l}|v|_{2,2,T}.$$

4.3.11. Assumption (A19). $\|L_h(I-I_h)v\|_{0,2,T} \leqslant \|L_h(I-I_{V_h})v\|_{0,2,T} + \|L_h(I_{V_h}-I_h)v\|_{0,2,T}.$ holds. The first term vanishes, since

$$L_hv(x_i) = v(x_i) = I_{V_h}v(x_i) = L_hI_{V_h}v(x_i).$$

To estimate the second term, we apply Assumption (A15) locally to see that

$$\|L_h(I_{V_h} - I_h)v\|_{0,2,T} \leqslant C\|(I_{V_h} - I_h)v\|_{0,2,T} \leqslant C\left[\|(I - I_{V_h})v\|_{0,2,T} + \|(I - I_h)v\|_{0,2,T}\right].$$

The first term in this relation can be estimated by means of [19, Lemma 2.14] (again with h_T^2 replaced by $\tilde{h}_T h_T$), whereas for the second one we use Assumption (A18) (*ii*).

In the final analysis, we have proved the following result.

Theorem 4.1. *Assume that a shape-regular family $\mathfrak{F}$ of partitions for P_1-parametric elements satisfies Assumption* (*A*8) *and that the family of triangulations generated by the subsimplices $\Omega_{ij}^{T,l}$ is locally quasi-uniform* (*see* (4.3)). *Then, the discrete form n_h has the following properties*:

semidefiniteness: $n_h(w_h, v_h, v_h) \geqslant 0$;

Lipschitz-continuity: $|n_h(w_h, u_h, v_h) - n_h(z_h, u_h, v_h)| \leqslant C\|w_h - z_h\|_h \|u_h\|_h \|v_h\|_h$;

onsistency: $|n(w, u, v_h) - n_h(I_h w, I_h u, v_h)| \leqslant C\tilde{h}\|w\|_{2,2,\Omega}\|u\|_{2,2,\Omega}\|v_h\|_h$,

$u_h, v_h, w_h \in V_h,\ u, w \in W_2^2(\Omega)^d \bigcap V$,

where $\tilde{h} := \max_{T\in\mathfrak{T}_h} \tilde{h}_T$ *and* $\tilde{h}_T$ *is defined according to (4.11).*

4.4. Error estimates. Now we return to the estimation of terms (2.11) – (2.19) in the error equation.

4.4.1. Estimation of (2.11) – (2.13). If the solution (u, p) satisfies the smoothness assumption (2.5), then we can apply Theorem 4.1, i.e., for $w := u$, $w_h := I_h u$ and arbitrary $v_h \in V_h$,

$$|n_h(w_h, w_h, v_h) - \tilde{n}_h(u, u, v_h)| \leqslant C\tilde{h}\|u\|_{2,2,\Omega}^2 \|v_h\|_h,$$

$$|n_h(u_h, u_h, v_h) - n_h(w_h, u_h, v_h)| \leqslant C\|u_h - w_h\|_h \|u_h\|_h \|v_h\|_h, \quad n_h(w_h, v_h, v_h) \geqslant 0$$

hold.

4.4.2. Estimation of (2.14). To estimate this term, we introduce the L_1-lumping operator $J_h :\ Q \to Q_h$ by

$$J_h q(x) := \frac{1}{\mathrm{meas}_d(T)} \int\limits_T q\,dx, \quad \forall x \in T \in \mathfrak{T}_h.$$

Then, by [19, p. 67], we get the desired result:

$$(p - p_h, \nabla\cdot v_h)_h = (p - J_h p, \nabla\cdot v_h)_h \leqslant Ch|p|_{1,2,\Omega}\|v_h\|_h.$$

4.4.3. Estimation of (2.15). By the definition (4.9), (4.10) of the discrete $L_2(\Omega)$-inner product $(\cdot,\cdot)_l$, we have

$$(\partial_t(u - w_h), v_h)_l = (L_h(I - I_h)\partial_t u, L_h v_h) \leqslant \|L_h(I - I_h)\partial_t u\|_{0,2,\Omega}\|L_h v_h\|_{0,2,\Omega},$$

where we have used the fact that the differentiation commutes with lumping and interpolation.

For $v \in W_p^1(\Omega)$, $p > d$, it is possible to prove the following estimate:

$$\|L_h(I - I_h)v\|_{0,2,T} \leqslant C\,h_T\|v\|_{1,p,T}, \quad T \in \mathfrak{T}_h, \tag{4.13}$$

where C is a positive constant independent of T and v. This can be done by a nearly word-for-word transfer of the arguments used in the above verification of Assumption (A19), including the necessary modifications of the corresponding passages in [19, Sect. 2.3].

Using (4.13) with $v = \partial_t u$ and (A15), we get

$$|(\partial_t(u - w_h), v_h)_l| \leqslant Ch|\partial_t u|_{1,p,\Omega}\|v_h\|_{0,2,\Omega}.$$

4.4.4. Estimation of (2.16). By the triangle inequality, there holds

$$|(\partial_t u, v_h)_l - (\partial_t u, v_h)| = |(L_h\partial_t u, L_h v_h) - (\partial_t u, v_h)| \leqslant |((L_h - I)\partial_t u, L_h v_h)| + |(\partial_t u, (L_h - I)v_h)|.$$

For the first term, we have

$$|((L_h - I)\partial_t u, L_h v_h)| \leqslant \|(L_h - I)\partial_t u\|_{0,2,\Omega}\|L_h v_h\|_{0,2,\Omega} \leqslant$$
$$\operatorname{meas}_d(\Omega)^{(p-2)/(2p)}\|(L_h - I)\partial_t u\|_{0,p,\Omega}\|L_h v_h\|_{0,2,\Omega}.$$

Now we use (A16) (i) and (A15) to get

$$|((L_h - I)\partial_t u, L_h v_h)| \leqslant Ch|\partial_t u|_{1,p,\Omega}\|v_h\|_{0,2,\Omega}.$$

To estimate the second term, we apply (A16)(ii). As a result of both estimates, we get

$$|(\partial_t u, v_h)_l - (\partial_t u, v_h)| \leqslant Ch\left[|\partial_t u|_{1,p,\Omega}\|v_h\|_{0,2,\Omega} + \|\partial_t u\|_{0,2,\Omega}\|v_h\|_h\right].$$

Since $\|\partial_t u\|_{0,2,\Omega} \leqslant C\|\partial_t u\|_{0,p,\Omega}$ and $\|v_h\|_{0,2,\Omega} \leqslant \|v_h\|_h$, we finally get

$$|(\partial_t u, v_h)_l - (\partial_t u, v_h)| \leqslant Ch\|\partial_t u\|_{1,p,\Omega}\|v_h\|_h.$$

4.4.5. Estimation of (2.17). For we have, by (A18)(ii),

$$|\varepsilon(\nabla(u - w_h), \nabla v_h)_h| \leqslant \varepsilon Ch\|u\|_{2,2,\Omega}|v_h|_h.$$

4.4.6. Estimation of (2.18). This estimate runs as in the estimation of (2.16) with $\partial_t u$ replaced by f, i.e.

$$|(f, v_h) - (f, v_h)_l| \leqslant Ch\|f\|_{1,p,\Omega}\|v_h\|_h.$$

4.4.7. Estimation of (2.19). The estimate (2.19) was proved in [19, Lemma 4.14 and Lemma 4.15 together with the remark on p. 47]:

$$|\langle\tilde{f}_h, v_h\rangle| \leqslant Ch[\|u\|_{2,2,\Omega} + \|p\|_{1,2,\Omega}]\|v_h\|_h.$$

4.4.8. Summary. Collecting all the above estimates of terms (2.11) – (2.19) together, we obtain with $v_h := w_h - u_h$, $w_h := I_h u$ the following inequality:

$$(\partial_t v_h, v_h)_l + \varepsilon(\nabla v_h, \nabla v_h)_h \leqslant \tilde{h}\Psi_1(u, p, f)\|v_h\|_h + \Psi_2(u_h)\|v_h\|_h^2, \tag{4.14}$$

where

$$\Psi_1(u, p, f) := C[(\varepsilon + 1 + \|u\|_{2,2,\Omega})\|u\|_{2,2,\Omega} + \|\partial_t u\|_{1,p,\Omega} + \|p\|_{1,2,\Omega} + \|f\|_{1,p,\Omega}], \quad \Psi_2(u_h) := C\|u_h\|_h$$

with constants $C > 0$ independent of u, p, f, u_h, v_h, h.

By (4.8) and Corollary 3.3, there holds

$$\|v_h\|_{0,2,\Omega} \leqslant C_P|v_h|_h \tag{4.15}$$

with some constant $C_P > 0$, and thus

$$\|v_h\|_h \leqslant \sqrt{1 + C_P^2}|v_h|_h. \tag{4.16}$$

Furthermore, Assumptions (A15) and (4.15) imply that there is a constant $\beta > 0$ independent of h such that with $\|v_h\|_{0,h} := \sqrt{(v_h, v_h)_l}$,

$$\|v_h\|_{0,h} \leqslant \beta|v_h|_h. \tag{4.17}$$

Because of the relation

$$(\partial_t v_h(t), v_h(t))_l = \frac{1}{2}\frac{d}{dt}(v_h(t), v_h(t))_l = \frac{1}{2}\frac{d}{dt}\|v_h(t)\|_{0,h}^2,$$

it follows from (4.14) and (4.16) that

$$\frac{1}{2}\frac{d}{dt}\|v_h(t)\|_{0,h}^2 + \varepsilon|v_h(t)|_h^2 \leqslant \tilde{h}\sqrt{1 + C_P^2}\Psi_1(u, p, f)|v_h|_h + (1 + C_P^2)\Psi_2(u_h)|v_h|_h^2 \leqslant$$

$$\tilde{h}^2\frac{1 + C_P^2}{\varepsilon}\Psi_1^2(u, p, f) + \left[\frac{\varepsilon}{4} + (1 + C_P^2)\Psi_2(u_h)\right]|v_h|_h^2. \tag{4.18}$$

If (4.17) is used to estimate one half of the term $\varepsilon|v_h(t)|_h^2$, a simple rearrangement of (4.18) leads to

$$\frac{d}{dt}\|v_h(t)\|_{0,h}^2 + \frac{\varepsilon}{\beta^2}\|v_h(t)\|_{0,h}^2 + \left[\frac{\varepsilon}{2} - 2(1 + C_P^2)\Psi_2(u_h)\right]|v_h|_h^2 \leqslant 2\tilde{h}^2\frac{1 + C_P^2}{\varepsilon}\Psi_1^2(u, p, f).$$

Multiplying this relation by $e^{\gamma t}$, where $\gamma := \varepsilon/\beta^2$, the identity

$$\frac{d}{dt}(e^{\gamma t}\|v_h(t)\|_{0,h}^2) = e^{\gamma t}\frac{d}{dt}\|v_h(t)\|_{0,h}^2 + \gamma e^{\gamma t}\|v_h(t)\|_{0,h}^2$$

leads to

$$\frac{d}{dt}(e^{\gamma t}\|v_h(t)\|_{0,h}^2) + \left[\frac{\varepsilon}{2} - 2(1 + C_P^2)\Psi_2(u_h)\right] e^{\gamma t}|v_h|_h^2 \leqslant 2\tilde{h}^2\frac{1 + C_P^2}{\varepsilon}\Psi_1^2(u, p, f)e^{\gamma t}.$$

If $\|u_h\|_h$ is sufficiently small so that the term in square brackets becomes nonnegative (the precise formulation of this assumption is given in (4.20) below), then

$$\frac{d}{dt}(e^{\gamma t}\|v_h(t)\|_{0,h}^2) \leqslant 2\tilde{h}^2\frac{1 + C_P^2}{\varepsilon}\Psi_1^2(u, p, f)e^{\gamma t},$$

and the integration over $(0, t)$ results in

$$e^{\gamma t}\|v_h(t)\|_{0,h}^2 - \|v_h(0)\|_{0,h}^2 \leqslant 2\tilde{h}^2\frac{1 + C_P^2}{\varepsilon}\int_0^t \Psi_1^2(u, p, f)\, e^{\gamma s}\, ds$$

for all $t \in (0, t_\infty)$. Multiplying this by $e^{-\gamma t}$, we get the relation

$$\|v_h(t)\|_{0,h}^2 \leqslant \|v_h(0)\|_{0,h}^2\, e^{-\gamma t} + 2\tilde{h}^2\frac{1 + C_P^2}{\varepsilon}\int_0^t \Psi_1^2(u, p, f)\, e^{-\gamma(t-s)}\, ds.$$

In the final analysis, we have proved the following result for the case where $\|u_h\|_h$ is sufficiently small.

Theorem 4.2. *Assume that the shape-regular family $\mathcal{F}$ of partitions for P_1-parametric elements satisfies Assumption* (*A*8) *and that the family of triangulations generated by the subsimplices $\Omega_{ij}^{T,l}$ is locally quasiuniform* (*see* (4.3)).

Further assume that the unique weak solution of (2.3) *with $u_0 \in W \bigcap V$ satisfies* (2.5), *i.e.,*

$$(u,p) \in L_2((0,t_\infty), W_2^2(\Omega)^d) \times L_2((0,t_\infty), W_2^1(\Omega)),$$

and, for some $p > d$, the condition

$$\int_0^t \left[(\varepsilon + 1 + \|u\|_{2,2,\Omega})\|u\|_{2,2,\Omega} + \|\partial_t u\|_{1,p,\Omega} + \|p\|_{1,2,\Omega} + \|f\|_{1,p,\Omega}\right]^2 e^{-\varepsilon(t-s)/\beta^2}\, ds < \infty, \quad (4.19)$$

is met, where β is the constant from (4.17).

Finally, let the semidiscrete problem (2.4) with $u_{0h} := I_h u_0$ have a unique solution $(u_h, p_h) \in V_h \times Q_h$ such that $\|u_h\|_h$ is sufficiently small in the following sense:

There exists a sufficiently small constant $c_0 > 0$ independent of h such that

$$\sup_{(0,t_\infty)} \|u_h\|_h \leqslant c_0 \varepsilon. \quad (4.20)$$

Then the following error estimate for the semidiscrete velocity field holds on $(0,t_\infty)$:

$$\|I_h u(t) - u_h(t)\|_{0,h} \leqslant \tilde{h}\sqrt{2\frac{1+C_P^2}{\varepsilon}}\, C_w,$$

where the quantity C_w is the square root of the left-hand side of (4.19).

5. Conclusion

In this paper we discussed a general framework for the finite-volume-based discretization of the nonlinear convective term in the incompressible Navier — Stokes equations. The proposed approach makes it possible to derive an estimate of the semidiscrete velocity error measured in a discrete L_2-norm without use of any linearized stability theory. As long as the numerical solution satisfies a certain smallness assumption, the stationary pendant of which is widely used (cf. [17, Sect. IV.2]), it has been shown that the constant in the error estimate is time-independent and of order $\mathcal{O}(\varepsilon^{-1/2}) = \mathcal{O}(\sqrt{Re})$, but not $\mathcal{O}(\exp(\varepsilon^{-1})) = \mathcal{O}(\exp(Re))$.

References

1. L. Angermann, *Error analysis of upwind-discretizations for the steady-state incompressible Navier-Stokes equations*, Preprint Nr. 33, Fakultät für Mathematik, Otto-von-Guericke-Universität Magdeburg, 1998.

2. L. Angermann, *Error analysis of upwind-discretizations for the steady-state incompressible Navier-Stokes equations*, Advances in Computational Mathematics, **13** (2000), pp. 167–198.

3. L. Angermann, *The one-step Θ-method for spatially stabilized finite volume discretizations of parabolic equations*, in: *Finite volumes for complex applications III — problems and perspectives* (R. Herbin and D. Kröner, eds.), Hermes Penton Science, London, 2002, pp. 25–39.

4. P. Angot, V. Dolejší, M. Feistauer, and J. Felcman, *Analysis of a combined barycentric finite volume — nonconforming finite element method for nonlinear convection-diffusion problems*, Appl. Math., **43** (1998), no. 4, pp. 263–310.

5. B. Bermúdez, A. Nikolás, and F. Sánchez, *On operator splitting methods with upwinding for the unsteady Navier — Stokes equations*, East-West J. Numer. Math., **4** (1996), no. 2, pp. 83–98.

6. S. Brenner and L. Scott, *The mathematical theory of finite element methods*, Springer-Verlag, New York–Berlin–Heidelberg, 2002, texts in Applied Mathematics, Vol. 15, 2nd ed.

7. L. Dechevski and E. Quack, *On the Bramble — Hilbert lemma*, Numer. Funct. Anal. Optimiz., **11** (1990), no. 5/6, pp. 485–495.

8. V. Dolejší, M. Feistauer, J. Felcman, and A. Kliková, *Error estimates for barycentric finite volumes combined with nonconforming finite elements applied to nonlinear convection-diffusion problems*, Appl. Math., **47** (2002), no. 4, pp. 301–340.

9. M. Feistauer, J. Felcman, M. Lukácová-Medvid'ova, and G. Warnecke, *Error estimates for a combined finite volume-finite element method for nonlinear convection-diffusion problems*, SIAM J. Numer. Anal., **36** (1999), no. 5, pp. 1528–1548.

10. M. Feistauer, J. Slavík, and P. Stupka, *On the convergence of a combined finite volume-finite element method for nonlinear convection-diffusion problems. Explicit schemes*, Numer. Meth. PDE, **15** (1999), no. 2, pp. 215–235.

11. T. Ikeda, *Maximum principle in finite element models for convection-diffusion phenomena*, North-Holland, Amsterdam–New York–Oxford, 1983.

12. H. Kanayama and K. Toshigami, *A partial upwind finite element approximation for the stationary Navier-Stokes equations*, Comput. Mech., **5** (1989), no. 2/3, pp. 209–216.

13. M. Marion and R. Temam, *Navier-Stokes equations: theory and approximation*, in: *Handbook of numerical analysis, Vol. VI*, North-Holland, Amsterdam, 1998, pp. 503–688.

14. J. Miller and S. Wang, *An exponentially fitted finite volume method for the numerical solution of 2D unsteady incompressible flow problems*, J. Comput. Phys., **115** (1994), no. 1, pp. 56–64.

15. S. Patankar, *Numerical heat transfer and flow*, Hemisphere Publishing Corporation, McGraw-Hill, New York, 1980.

16. S. Perron, S. Boivin, and J.-M. Hérard, *A finite volume method to solve the 3D Navier-Stokes equations on unstructured collocated meshes*, Comput. & Fluids, **33** (2004), no. 10, pp. 1305–1333.

17. H.-G. Roos, M. Stynes, and L. Tobiska, *Numerical methods for singularly perturbed differential equations. Convection-diffusion and flow problems*, Springer-Verlag, Berlin–Heidelberg–New York, 1996, springer series in computational mathematics, vol. 24.

18. F. Schieweck, *On the order of two nonconforming finite element approximations of upwind type for the Navier-Stokes equations*, in: *Numerical methods for Navier-Stokes equations* (F.-K. Hebeker, R. Rannacher, and G. Wittum, eds.), Vieweg, Braunschweig-Wiesbaden, 1994, pp. 249–258, notes on Numerical Fluid Mechanics, vol. 47.

19. F. Schieweck, *Parallele Lösung der stationären inkompressiblen Navier — Stokes Gleichungen*, Habilitationsschrift, Fakultät für Mathematik, Universität Magdeburg, 1997.

20. F. Schieweck and L. Tobiska, *A nonconforming finite element method of upstream type applied to the stationary Navier-Stokes equations*, RAIRO Modél. Math. Anal. Numér., **23** (1989), pp. 627–647.

21. L. Tobiska, *A three-dimensional nonconforming finite element method of upstream type and its application to the Navier — Stokes equations*, in: *Numerical methods in singularly perturbed problems* (H.-G. Roos, A. Felgenhauer, and L. Angermann, eds.), TU Dresden, 1991, pp. 155–164.