# TECHNISCHE UNIVERSITÄT BERLIN

# Finite Element Decomposition and Minimal Extension for Flow Equations

**Robert Altmann** **Jan Heiland**

**Preprint 2013/11**

**Preprint-Reihe des Instituts für Mathematik**
**Technische Universität Berlin**
http://www.math.tu-berlin.de/preprints

**Preprint 2013/11** **April 2013**

# FINITE ELEMENT DECOMPOSITION AND MINIMAL EXTENSION FOR FLOW EQUATIONS

ROBERT ALTMANN, JAN HEILAND

Abstract. In the simulation of flows, the correct treatment of the pressure variable is the key to stable time-integration schemes. This paper contributes a new approach based on the theory of differential-algebraic equations. Motivated by the index reduction technique of minimal extension, a decomposition of finite element spaces is proposed that ensures stable and accurate approximations.

The presented decomposition – for standard finite element spaces used in CFD – preserves sparsity and does not call on variable transformations which might change the meaning of the variables. Since the method is eventually an index reduction, high index effects leading to instabilities are eliminated. As a result, all constraints are maintained and one can apply semi-explicit time integration schemes.

**Key words.** Navier-Stokes equations, time integration schemes, finite element method, index reduction, operator DAE.

**AMS subject classifications. 76M10** (Finite element methods for fluids), **65L80** (Methods for DAEs), **65J10** (Abstract evolution equations).

## 1. Introduction

The spatial semi-discretization of Navier-Stokes equations (NSE) leads to differential-algebraic equations (DAE) of differentiation index 2, cf. [41], and takes the form

$$M\dot{u} + K(u) - B^T p = f, \quad u(0) = a \tag{1.1a}$$

$$Bu = 0 \tag{1.1b}$$

For time integration schemes one has to take care of the differential-algebraic structure that require implicit schemes and that can cause a reduction of the convergence order up to possible divergence [17, 25]. To avoid divergence for low-order schemes, a general approach is the reformulation of the equations as an equivalent or arbitrarily close system of index 1 [17, 25, 41].

For the semi-discrete NSE a variety of methods has been developed that can be roughly classified into penalty methods [22, 33], pressure correction

The work of R. Altmann was supported by the ERC Advanced Grant "Modeling, Simulation and Control of Multi-Physics Systems" MODSIMCONMP and the Berlin Mathematical School BMS.

The work of J. Heiland was funded by a scholarship by "Studienstiftung des deutschen Volkes".

or projection methods [17], and divergence-free methods. These solution procedures used for the numerical time integration can be interpreted as an index reduction [41].

In penalization methods, one adds a term $\lambda^{-1}p$, $\lambda \gg 1$, to the left hand side of (1.1b) and obtains an ODE for $u$ via $M\dot{u} + K(u) - \lambda B^T B u = f$. Disadvantages of this approach are the stiffness of the obtained system, time step restrictions [33], and that the pressure has to be computed separately.

In projection methods, also referred to as operator splitting or pressure correction methods, one uses a guess for the pressure to compute an approximate velocity update $\tilde{u}$ via (1.1a) in every time step. Then, one computes the components $\tilde{u} = u_0 + u_\perp$, with $u_0$ satisfying $Bu_0 = 0$ and $u_\perp$ being in the span of $B^T$. This splitting or projection requires the solution of the so-called pressure Poisson equation that also defines the update for the pressure $p$. The main disadvantage of this approach is the requirement of boundary conditions for the pressure update that have no physical meaning [37].

The two mentioned approaches decouple the pressure and velocity computation. This is computationally beneficial since (1.1a) decomposes into two smaller systems. As elaborated in [41], this decoupling is incomplete and depends strongly on the heuristic penalization parameter or the time step.

A complete decoupling is obtained in divergence-free formulations that reduce (1.1a) to an ODE for the divergence-free components of $u$. The presence of a divergence-free basis for $u$ is optimal for the approximation of the velocity since the system is reduced to a subspace of the velocity space and the constraints (1.1b) are fulfilled a priori. An overview of divergence-free elements is provided in [17, Ch. 3.13.7]. However, these elements are rarely used in simulations because of their difficult implementation [17, Ch. 3.12.2].

Note that the use of quasi divergence-free elements, see e.g. [29], reduces the systems size but leaves the DAE structure unchanged.

One can also use discrete realizations of the Leray or Helmholtz projector $\pi$ [16, Cor. I.3.4], that projects $u$ onto the kernel of $B$. Applying $\pi^T$ to (1.1a) and using that $\pi^T M = M\pi$, one obtains

$$M\frac{d}{dt}(\pi u) + \pi^T K(\pi u) = \pi^T f + \begin{pmatrix} \pi^T B^T p \end{pmatrix}, \quad \pi u(0) = \pi a \tag{1.2}$$

which is an ODE for $\pi u$, see e.g. [21]. One difficulty of this approach is that the direct use of $\pi$ is not feasible for large systems [21], while an approximate realization of $\pi$ giving $\pi^T B^T \approx \varepsilon I \neq 0$ introduces a systematic error to the solution. Additionally, the numerical integration of (1.2) leads to a drift from the constraint manifold and requires stabilization [6].

Also, one may resolve the algebraic constraints numerically, e.g. via a QR-decomposition of $B$. This approach is again not taken in practice because the variable transformation $u \leftarrow Q\tilde{u}$ is computationally unfeasible already for moderately sized systems. Another disadvantage, and this holds also for divergence-free elements, is that the associated equation for the pressure is ill-conditioned, as demonstrated in Section 3.1 below.

Recent approaches [26] for the numerical construction of sparse and well-conditioned divergence-free bases, i.e., a null space of $B$, only tackle the problem of infeasibility.

Further common methods in computational fluid dynamics (CFD) that use heuristic strategies for stabilization are beyond our consideration of generally, algebraically and consistently derived approaches.

We propose a variant of minimal extension [24, 25] for an index reduction of the abstract equations and develop efficient algorithms for particular but commonly used finite element discretizations. Recently, minimal extension has been formulated for elastodynamics [1] in an abstract setting, consistent in terms of the differential-algebraic structure with finite element discretizations.

This new approach can be seen as resolving the algebraic constraints while at the same time avoiding the mentioned above difficulties. The variables are transformed via a simple permutation and thus keep their physical meaning. The sole application of a permutation preserves sparsity and is well conditioned. Since the so-called hidden constraint $B\dot{u} = 0$ is added to the system instead of implicitly eliminated, instabilities are reduced. Finally, the pressure $p$ remains a physically valid part of the system, rather than being eliminated or functioning as a velocity correction.

The increase of the system size is compensated by the direct applicability of efficient time stepping schemes.

This paper is organized as follows. In Section 2 the unsteady NSE is formulated as constrained operator differential equation. Following the ideas of minimal extension, we reformulate the so-called operator DAE such that a spatial discretization leads to a DAE of differentiation index 1. This property requires certain assumptions on the finite element spaces which are presented in Section 3. In particular, a splitting of the velocity ansatz space is necessary. We show the advantages of this method and give examples for such splittings for standard discretization schemes such as Crouzeix-Raviart [14] and Taylor-Hood [35].

In Section 4 we present the benefits of the presented approach for numerical time integration by means of a non-viscous two-dimensional flow equation.

## 2. Operator Formulation

We consider the unsteady NSE on a domain $\Omega \subset \mathbb{R}^n$, $n \in \{2, 3\}$ with twice differentiable boundary $\partial\Omega$ in a time interval $(0, T)$,

$$\dot{u} + (u \cdot \nabla)u - \frac{1}{\mathrm{Re}} \triangle u + \nabla p = \beta \qquad \text{in } \Omega \times (0, T), \tag{2.1a}$$

$$\operatorname{div} u = 0 \qquad \text{in } \Omega \times (0, T), \tag{2.1b}$$

$$u = 0 \qquad \text{on } \partial\Omega \times (0, T), \tag{2.1c}$$

$$u(\cdot, 0) = a. \tag{2.1d}$$

This system describes the evolution of the velocity field $u(t) \in (\Omega \to \mathbb{R}^n)$ and the pressure $p(t) \in (\Omega \to \mathbb{R})$ for a given parameter $\mathrm{Re} > 0$, an initial value $a$ and a volume force $\beta$.

2.1. **Preliminaries.** For the basic definition of Sobolev spaces on a domain $\Omega$, as the space of square integrable functions $L^2(\Omega)$ that possess a weak derivative in $L^2(\Omega)$, its subspace $H_0^1(\Omega)$ of functions that are weakly differentiable and zero on the boundary $\partial\Omega$, and of Bochner spaces like $L^2(0,T;L^2(\Omega))$ or $H^1(0,T;L^2(\Omega)$, we refer the reader to [34].

To shorten notation, we define the spaces

$$\mathcal{V} := [H_0^1(\Omega)]^n, \quad \mathcal{H} := [L^2(\Omega)]^n, \quad \text{and } \mathcal{Q} := L^2(\Omega)/\mathbb{R}.$$

The space $\mathcal{V}$ is densely and continuously embedded in $\mathcal{H}$ and thus, the identification of $\mathcal{H}$ with its dual $\mathcal{H}'$ via the Riesz isomorphism gives the evolution triple $\mathcal{V} \subset \mathcal{H} \subset \mathcal{V}'$. Let $\mathcal{W}^{2,1}(0,T)$ be the space of functions $u \in L^2(0,T;\mathcal{V})$, with weak time derivative $\dot{u} \in L^1(0,T;\mathcal{V}')$.

We consider a weak formulation of (2.1): Given right-hand sides $\mathcal{F} \in L^2(0,T;\mathcal{V}')$, $\mathcal{G} \in L^2(0,T;\mathcal{Q}')$ and an initial condition $a \in \mathcal{H}$, we seek for $(u,p) \in \mathcal{W}^{2,1}(0,T) \times L^2(0,T;\mathcal{Q})$ satisfying

$$\dot{u}(t) + \mathcal{K}(u(t)) - \mathcal{B}'p(t) = \mathcal{F}(t) \qquad \text{in } \mathcal{V}', \tag{2.2a}$$

$$\mathcal{B}u(t) \qquad\qquad = \mathcal{G}(t) \qquad \text{in } \mathcal{Q}', \tag{2.2b}$$

$$u(0) = a \qquad \text{in } \mathcal{H}, \tag{2.2c}$$

a. e. on $(0,T)$. Because of the differential-algebraic structure in an abstract setting, we call (2.2) an *operator DAE*. Therein, the operators $\mathcal{K}\colon \mathcal{V} \to \mathcal{V}'$ and $\mathcal{B}\colon \mathcal{V} \to \mathcal{Q}'$ are defined via

$$\langle \mathcal{K}(u), v \rangle = \int_\Omega (u \cdot \nabla) u \cdot v \, \mathrm{d}x + \frac{1}{\mathrm{Re}} \int_\Omega \nabla u \cdot \nabla v \, \mathrm{d}x \tag{2.3}$$

and

$$\langle \mathcal{B}u, q \rangle = \int_\Omega (\operatorname{div} u) q \, \mathrm{d}x = \langle u, \mathcal{B}'q \rangle, \tag{2.4}$$

respectively, given $u \in \mathcal{V}$ and for all $v \in \mathcal{V}$ and $q \in \mathcal{Q}$.

Since $\mathcal{B}$ is bounded, the ansatz space $\mathcal{V}$ can be decomposed into the divergence-free space $\mathcal{V}_{\mathrm{df}}$ and its orthogonal complement $\mathcal{V}_{\mathrm{df}}^\perp$ with respect to the inner product of $\mathcal{V}$, i.e.,

$$\mathcal{V}_{\mathrm{df}} := \ker \mathcal{B} = \{u \in \mathcal{V} \mid \operatorname{div} u = 0\}, \qquad \mathcal{V} = \mathcal{V}_{\mathrm{df}} \oplus \mathcal{V}_{\mathrm{df}}^\perp. \tag{2.5}$$

This implies a unique decomposition of $u \in \mathcal{V}$ into $u = u_1 + u_2$ with $u_1 \in \mathcal{V}_{\mathrm{df}}$ and $u_2 \in \mathcal{V}_{\mathrm{df}}^\perp$.

2.2. **Existence of Solutions.** Classical existence results consider (2.2) with $\mathcal{G}(t) = 0$ formulated on the subspace of divergence-free functions $\mathcal{V}_{\mathrm{df}} \subset \mathcal{V}$. The problem then turns to: find $u_1 \in L^2(0,T;\mathcal{V}_{\mathrm{df}})$ with $\dot{u}_1 \in L^1(0,T;\mathcal{V}'_{\mathrm{df}})$ satisfying

$$\dot{u}_1(t) + \mathcal{K}_1(u_1(t)) = \mathcal{F}_1(t) \quad \text{in } \mathcal{V}'_{\mathrm{df}}, \tag{2.6a}$$

$$u_1(0) = a_1 \quad \text{in } \mathcal{H}, \tag{2.6b}$$

a. e. on $(0,T)$. Therein, let $\mathcal{F}_1 \in L^2(0,T;\mathcal{V}'_{\mathrm{df}})$, $a_1$ be in the closure of $\mathcal{V}_{\mathrm{df}}$ w.r.t. the norm of $\mathcal{H}$, and $\mathcal{K}_1 \colon \mathcal{V}_{\mathrm{df}} \to \mathcal{V}'_{\mathrm{df}}$ be defined as in (2.3). The formulation via divergence-free functions, in particular, eliminates the pressure from the equations.

There exists a solution $u_1 \in L^2(0,T;\mathcal{V}_{\mathrm{df}})$ satisfying (2.6), see [36, Thm. III.3.1], which is unique in the two-dimensional case [36, Thm. III.3.2]. Given $u$, one can generally establish a corresponding pressure $p$ as a distribution on $(0,T)\times\Omega$, cf. [36, pp. 307]. However, the pair $(u_1,p)$ only solves (2.2) under additional regularity conditions, cf. [31], and if $a = a_1$. In particular, if the values in (2.2a) are in $\mathcal{H}$ rather than in $\mathcal{V}'$, then system (2.2) can be split via the Helmholtz decomposition [16, Cor. I.3.4] into a part defining $u_1 \in L^2(0,T;\mathcal{V}_{\mathrm{df}})$ and a part that uniquely defines $p \in L^2(0,T;Q)$. This additional regularity is given globally in 2D and locally in time in 3D, if $\mathcal{F} \in L^2(0,T;\mathcal{H})$ and $a \in \mathcal{V}_{\mathrm{df}}$, cf. [34, Lems. 25.1,25.2]. Since a solution $u_1$ to (2.2) always solves (2.6), in 2D, it is unique.

An inhomogeneity $\mathcal{G}$ in the constraint (2.2b) is likely to appear in discretized schemes and for more general boundary conditions. For maximal generality, we will consider it present, as it imposes restrictions on the solvability of the equations.

Since $\mathcal{B} : \mathcal{V} \to \mathcal{Q}'$ has a right-inverse $\mathcal{B}^-$, see [16, Lem. I.4.1], the complement to $u_1$ is eventually given by $u_2 = \mathcal{B}^-\mathcal{G}$. Plugging this relation into (2.2), we obtain the remainder system

$$\dot{u}_1(t) + \mathcal{K}\big(u_1(t) + \mathcal{B}^-\mathcal{G}(t)\big) - \mathcal{B}'p(t) = \mathcal{F}(t) - \mathcal{B}^-\dot{\mathcal{G}}(t) \quad \text{in } \mathcal{V}', \tag{2.7a}$$

$$\mathcal{B}u_1(t) = 0 \quad \text{in } \mathcal{Q}', \tag{2.7b}$$

$$u_1(0) = a - \mathcal{B}^-\mathcal{G}(0) \quad \text{in } \mathcal{H}, \tag{2.7c}$$

which is well-posed, only if $\dot{\mathcal{G}}$ is at least in $L^1(0,T;\mathcal{Q}')$. Then, solvability for (2.7) can be established analogously to (2.2).

2.3. **Index Reduction.** The spatial discretization of (2.2) leads to a DAE of index 2. Here we use the concept of the perturbation index [18], that for semi-explicit systems as (2.2) and (2.8) coincides with the differentiation index [12]. Thus, a system is of index $d$ if $d$ is the smallest integer such that a perturbation of the right hand side $\delta$ causes a deviation in the solution that can be bounded via the first $d-1$ time derivatives of $\delta$, cf. [18, Def. 1.1]. In numerical simulations, the occurrence of derivatives of perturbations appears as divisions by powers of the small discretization parameters [18, p. 1].

Thus, it may be preferable to use equivalent formulations of lower index. The semi-explicit structure of the NSE allows for minimal extension [24], which reduces the index without transforming the variables. This is done by adding the time derivatives of the constraints, what leads to an overdetermined system, and then introducing a minimal number of variables to make the system square again.

Following this idea, we reformulate the operator DAE (2.2) to index-1 form. By this we mean that a certain discretization in space leads to a DAE of index 1. Using (2.5), we seek in equations (2.2) for $u_1$ and $u_2$ instead of $u$. The corresponding ansatz spaces read

$$u_1 \in \mathcal{W}^{2,1}(0,T) \cap L^2(0,T;\mathcal{V}_{\mathrm{df}}), \quad u_2 \in \mathcal{W}^{2,1}(0,T) \cap L^2(0,T;\mathcal{V}_{\mathrm{df}}^{\perp}).$$

Assuming sufficient regularity, we add the derivative of the constraint, the *hidden constraint*. Since the operator $\mathcal{B}$ is independent of time, the hidden constraint reads

$$\mathcal{B}\dot{u}_2(t) = \mathcal{B}\big[\dot{u}_1(t) + \dot{u}_2(t)\big] = \dot{\mathcal{G}}(t).$$

As a second step, we introduce a new variable $\tilde{u}_2 := \dot{u}_2$. The reformulated and extended problem then reads: find $u_1 \in H^1(0,T;\mathcal{V}_{\mathrm{df}})$, $u_2, \tilde{u}_2 \in L^2(0,T;\mathcal{V}_{\mathrm{df}}^{\perp})$, and $p \in L^2(0,T;\mathcal{Q})$ such that

$$\dot{u}_1(t) + \tilde{u}_2(t) + \mathcal{K}\big(u_1(t) + u_2(t)\big) - \mathcal{B}'p(t) = \mathcal{F}(t) \qquad \text{in } \mathcal{V}', \tag{2.8a}$$

$$\mathcal{B}u_2(t) = \mathcal{G}(t) \qquad \text{in } \mathcal{Q}', \tag{2.8b}$$

$$\mathcal{B}\tilde{u}_2(t) = \dot{\mathcal{G}}(t) \qquad \text{in } \mathcal{Q}', \tag{2.8c}$$

$$u_1(0) = a_1 \qquad \text{in } \mathcal{H}. \tag{2.8d}$$

*Remark* 2.1. Since $\mathcal{B}$ is constant in time, equations (2.8b-c) imply that $u_2$ is eventually in $H^1(0,T;\mathcal{V}_{\mathrm{df}}^{\perp})$.

*Remark* 2.2. For the spatial discretization in Section 3, we add to (2.8b) and to (2.8c) the vanishing terms $\mathcal{B}u_1(t)$ and $\mathcal{B}\dot{u}_1(t)$, respectively. This is necessary since we will deal with nonconforming finite elements, where the discrete version of $u_1$ does not vanish under the action of $\mathcal{B}$.

The derivation of system (2.8) proves the following theorem.

**Theorem 2.1.** *If $\mathcal{G} \in H^1(0,T;\mathcal{Q}')$ and $a = a_1 + \mathcal{B}^{-}\mathcal{G}(0)$, then the operator DAEs* (2.2) *and* (2.8) *have the same solution set.*

For the proof that the extended operator DAE (2.8) leads to an index-1 DAE, we refer to Section 3.2.

## 3. Discrete Formulation

This section is devoted to the spatial and temporal discretization of the NSE in two space dimensions. We show that the introduced splitting of the ansatz spaces provides an efficient simulation procedure.

We consider spatial discretizations by finite elements, i.e., we construct finite dimensional subspaces $V_h$ and $Q_h$ of $\mathcal{V}$ and $\mathcal{Q}$, respectively, based on

a triangulation $\mathcal{T}$ of the polygonal Lipschitz domain $\Omega$. The triangulation is assumed to be regular in the sense of Ciarlet [13]. Furthermore, we take for granted that the triangulation is shape regular [9, Ch. II.5]. In the sequel, $\mathcal{N}$ denotes the set of vertices of $\mathcal{T}$ and $\mathcal{E}$ set set of edges. The latter consists of interior, namely $\mathcal{E}_{\text{int}}$, and boundary edges, namely $\mathcal{E}_{\text{ext}}$. We focus on triangular meshes but will comment on quadrilateral elements in Section 3.6.

The finite dimensional approximation of the velocity $u(t)$ is given by the coefficient vector $q(t) \in \mathbb{R}^n$, which corresponds to a function in $V_h$. The discrete representative of the pressure is again denoted by $p(t) \in \mathbb{R}^m$. The semi-discretized version of system (2.2) reads

$$M\dot{q}(t) + K(q(t)) - B^T p(t) = f(t), \tag{3.1a}$$

$$Bq(t) = g(t). \tag{3.1b}$$

Therein, for a given basis $\{\Psi_j\}$ of $V_h$ and $\{\varphi_i\}$ of $Q_h$,

$$M = [m_{jk}] \in \mathbb{R}^{n\times n}, \quad m_{jk} := \int_\Omega \Psi_j \cdot \Psi_k \, \mathrm{d}x \tag{3.2a}$$

denotes the positive definite mass matrix and the nonlinear function $K$ is the discrete version of the operator $\mathcal{K}$,

$$K_j(q(t)) := \int_\Omega (q(t)\cdot\nabla)q(t)\cdot\Psi_j \, \mathrm{d}x + \frac{1}{\mathrm{Re}}\int_\Omega \nabla q(t)\cdot\nabla\Psi_j \, \mathrm{d}x, \tag{3.2b}$$

where we have assigned $q(t)$ with its function representation in $V_h$. The matrix $B = [b_{ij}] \in \mathbb{R}^{m\times n}$ is defined via

$$b_{ij} = \int_\Omega \varphi_i \operatorname{div} \Psi_j \, \mathrm{d}x. \tag{3.2c}$$

In the next subsection, we recall solution strategies of solving system (3.1) with the help of the QR algorithm and divergence-free finite elements. Afterwards, we propose a different ansatz which is based on the index-1 formulation which arises from the discretization of the operator DAE (2.8). This includes a decomposition of the finite element space $V_h$.

### 3.1. QR Decomposition and Divergence-free Elements.

For completeness, we address the case of eliminating all algebraic constraints and reducing the system to the so-called *inherent ODE*. In other words, we consider here the index-0 formulation of the NSE.

Numerically, this can be achieved by a QR decomposition of $B = \begin{bmatrix} 0 & R \end{bmatrix} Q$, with $R$ invertible and $Q$ unitarian. With the transformation $q =: Q^T\tilde{q}$ and the splitting $\tilde{q} = \begin{bmatrix} \tilde{q}_2^T & \tilde{q}_2^T \end{bmatrix}^T$, the divergence constraint (1.1a) becomes $\tilde{q}_2 = -R^{-1}g$. Then, a scaling of the momentum equation (1.1a) by $Q^T$ gives the decoupled system

$$\tilde{M}_{11}\dot{\tilde{q}}_1 + \tilde{K}_{11}(\tilde{q}_1) = \tilde{f}_1, \tag{3.3a}$$

$$-R^T p = -\tilde{M}_{21}\dot{\tilde{q}}_1 - \tilde{K}_{21}(\tilde{q}_1) + \tilde{f}_2. \tag{3.3b}$$

The subscripts refer to the considered parts of the equations and the tilde denotes the coefficients after the transformation of the system and the substitution of $\tilde{q}_2$ by $-R^{-1}g$.

Since $\tilde{M}_{11}$ is invertible, (3.3a) is equivalent to a standard ODE for $\tilde{q}_1$. Thus, one can expect stable approximations to $q = Q^T\tilde{q}$. However, the pressure $p$ as defined by (3.3b) requires $\dot{\tilde{q}}_1$ and $\tilde{K}_{21}(\tilde{q}_1)$, i.e., discrete time and space derivatives. In a numerical realization, this amplifies a non-smooth error in $\tilde{q}_1$ by $\tau^{-1}$ or $h^{-2}$, where $\tau$ and $h$ are length scales of time and space discretization, respectively.

Any such decomposition, not excluding divergence-free elements, would suffer from these instabilities in the pressure approximation for $\tau$ and $h$ tending to zero. Apart from that, divergence-free finite element discretizations [11, p. 267] are rarely used because of their cost, i.e., the high number of degrees of freedom and the lack of simplicity.

3.2. **Index-1 Formulation.** As announced above, in this subsection we show that a proper semi-discretization in space of system (2.8) leads to a DAE of index 1. Since $\mathcal{V}$ was decomposed in Section 2.3 into $\mathcal{V}_{\mathrm{df}} \oplus \mathcal{V}_{\mathrm{df}}^{\perp}$, we also decompose the finite dimensional space $V_h$. We denote the approximation space of $\mathcal{V}_{\mathrm{df}}$ by $V_{h,1}$, its complement $\mathcal{V}_{\mathrm{df}}^{\perp}$ is discretized by $V_{h,2}$. Furthermore, we assume that the direct sum of $V_{h,1}$ and $V_{h,2}$ is again $V_h$. Note that we do not assume the discretization to be conform, i.e., we allow for $V_{h,1} \not\subset \mathcal{V}_{\mathrm{df}}$ and $V_{h,2} \not\subset \mathcal{V}_{\mathrm{df}}^{\perp}$, although $V_h \subset \mathcal{V}$.

With $q_1$, $q_2$, and $\tilde{q}_2$ denoting the semi-discrete approximations of $u_1$, $u_2$, and $\tilde{u}_2$, respectively, we yield the spatial discretization of system (2.8) which reads

$$M \begin{bmatrix} \dot{q}_1(t) \\ \dot{\tilde{q}}_2(t) \end{bmatrix} + K\Big( \begin{bmatrix} q_1(t) \\ q_2(t) \end{bmatrix} \Big) - B^T p(t) = f(t), \tag{3.4a}$$

$$B \begin{bmatrix} q_1(t) \\ q_2(t) \end{bmatrix} = g(t), \tag{3.4b}$$

$$B \begin{bmatrix} \dot{q}_1(t) \\ \dot{\tilde{q}}_2(t) \end{bmatrix} = \dot{g}(t) \tag{3.4c}$$

with $M$, $K$, and $B$ as defined in (3.2). Here, we assume the basis of $V_h$ to be ordered according to its decomposition into $V_{h,1}$ and $V_{h,2}$. In the sequel, we analyse for which discretizations the DAE (3.4) is of index 1.

We require the standard stability condition for the spatial discretization [16, Ch. II] that reads: There exists a positive constant $c \in \mathbb{R}$ such that

$$\inf_{q_h \in Q_h} \sup_{v_h \in V_h} \frac{\langle \mathcal{B} v_h, q_h \rangle}{\|v_h\|_{\mathcal{V}} \|q_h\|_{\mathcal{Q}}} \geq c > 0. \tag{3.5}$$

From (3.5) we infer that $B$ has full row rank and that there is a decomposition $V_h = V_{h,1} \oplus V_{h,2}$ such that the submatrix of the columns accounting for $V_{h,2}$ is invertible. Formally, we put this into the following assumption.

*Assumption* 3.1. The finite element spaces $V_{h,1}$, $V_{h,2}$, and $Q_h$ satisfy: The matrix representation $B$ as defined in (3.2c) has the block structure $B = [B_1 \; B_2]$ with a nonsingular square matrix $B_2$ that contains the columns corresponding to $V_{h,2}$.

As a direct consequence of Assumption 3.1, we have that $\dim V_{h,2} = \dim Q_h$. In Section 3.5 we give examples how to decompose $V_h$ for certain finite element spaces used in CFD to meet Assumption 3.1.

**Theorem 3.1.** *Every finite element discretization of* (2.8) *with spaces* $V_{h,1}$, $V_{h,2}$, *and* $Q_h$ *satisfying Assumption 3.1 leads to a DAE of index 1.*

*Proof.* We follow the proof of [25, Th. 6.12] and show that under Assumption 3.1 the DAE (3.4) has index 1. A multiplication of (3.4a) by $BM^{-1}$ from the left and the relation (3.4c) give

$$-BM^{-1}B^T p = BM^{-1}f - BM^{-1}K\Big(\begin{bmatrix} q_1 \\ q_2 \end{bmatrix}\Big) - \dot{g}. \tag{3.6}$$

Since $M$ is positive definite and $B$ is of full rank by Assumption 3.1, $BM^{-1}B^T$ is invertible. Thus, we can express the pressure $p$ in terms of $f, \dot{g}, q_1$, and $q_2$. By Assumption 3.1 and (3.4b), we have

$$q_2 = B_2^{-1}g - B_2^{-1}B_1 q_1. \tag{3.7}$$

Finally, if we insert (3.6) and (3.7) into equation (3.4a), we obtain

$$M\begin{bmatrix} \dot{q}_1 \\ \tilde{q}_2 \end{bmatrix} = f + B^T p - K\Big(\begin{bmatrix} q_1 \\ q_2 \end{bmatrix}\Big) =: f^*(f, g, \dot{g}, q_1).$$

Since $M$ is invertible, this provides an ODE in $q_1$. Thus, we can solve for $q_1$, $q_2$ (by (3.7)), $\tilde{q}_2$ (by (3.4c)), and $p$ (by (3.6)), i.e. the DAE (3.4) is of index 1. □

*Remark* 3.1. The inf-sup condition (3.5) ensures a bound on the inverse of $BM^{-1}B^T$ from equation (3.6) independent of the discretization parameter $h$ and thus, stability in the spatial approximation of the pressure.

*Remark* 3.2. The reordering of the basis of $V_h$, that ensures Assumption 3.1, always exists – if (3.5) holds – and is basically a permutation of the velocity variables. This time-independent transformation is applied and inverted in exact arithmetics and preserves sparsity of the coefficient matrices.

3.3. **Time Integration.** The semi-discretized NSE (3.1) represents a semi-explicit index-2 DAE. For these systems, implicit time-stepping schemes such as the *Radau IIa* or *backward differencing* methods provide stable approximations of arbitrary order, provided that the inhomogeneities are sufficiently smooth [19]. These methods, however, require the solution of the full coupled nonlinear system at every stage. A compromise of the stability of implicit and the low computational load of explicit schemes is given by half-explicit schemes, that are explicit in the dynamic equations and implicit in the algebraic part.

Half-explicit Runge-Kutta (RK) methods were investigated for index-1 DAEs in [5, 27]. Methods for the index-2 case are provided e.g. in [4, 19]. Generally speaking, the application to index-1 problems is straight-forward while index-2 problems require specific treatments and possibly additional stages in the RK method.

We illustrate the different behavior with respect to inaccuracies of the index-1 and index-2 formulation of the NSE, using an explicit Euler method [17, Ch. 3.16.1] for the dynamical part. Superscripts $+, c$, and $-$ denote the next, current, and previous iterates, respectively. For the index-2 equation (3.1), the update to $(q^+, p^c)$ from the current iterate $(q^c, p^-)$ via a time step of length $\tau$ is obtained via

$$\begin{bmatrix} \frac{1}{\tau}M & -B^T \\ B & 0 \end{bmatrix} \begin{bmatrix} q^+ \\ p^c \end{bmatrix} = \begin{bmatrix} \frac{1}{\tau}Mq^c + f^c - K(q^c) \\ g^+ \end{bmatrix}. \tag{3.8}$$

For the update of the index-1 formulation (3.4) we propose the solution of

$$\begin{bmatrix} \frac{1}{\tau}M_{11} & M_{12} & -B_1^T & 0 \\ \frac{1}{\tau}M_{21} & M_{22} & -B_2^T & 0 \\ \frac{1}{\tau}B_1 & B_2 & 0 & 0 \\ B_1 & 0 & 0 & B_2 \end{bmatrix} \begin{bmatrix} q_1^+ \\ \tilde{q}_2^c \\ p^c \\ q_2^+ \end{bmatrix} = \begin{bmatrix} \frac{1}{\tau}M_{11}q_1^c + f_1^c - K_1(q_1^c, q_2^c) \\ \frac{1}{\tau}M_{21}q_1^c + f_2^c - K_2(q_1^c, q_2^c) \\ \frac{1}{\tau}B_1 q_1^c + \dot{g}^c \\ g^+ \end{bmatrix}. \tag{3.9}$$

The different stability properties become evident, if one examines the inherent equation for the pressure update $p^c$, derived via premultiplying the upper part of the equations by $BM^{-1}$. In the index-2 case (3.8), this leads to

$$-BM^{-1}B^T p^c = \frac{Bq^c - Bq^+}{\tau} + BM^{-1}[f^c - K(q^c)]. \tag{3.10}$$

The index-1 formulation yields for the pressure

$$-BM^{-1}B^T p^c = \frac{1}{\tau}B \begin{bmatrix} q_1^c - q_1^+ \\ -\tau\tilde{q}_2^c \end{bmatrix} + BM^{-1}[f^c - K(q_1^c, q_2^c)]. \tag{3.11}$$

If the equations are solved up to a residual of size $\varepsilon^c$, the dominating difference in the pressure definition in the index-1 and index-2 formulation lies in the terms

$$\frac{Bq^c - Bq^+}{\tau} = \frac{g^c - g^+}{\tau} + \frac{\varepsilon^c - \varepsilon^+}{\tau} \quad \text{and} \quad \frac{1}{\tau}B \begin{bmatrix} q_1^c - q_1^+ \\ -\tau\tilde{q}_2^c \end{bmatrix} = -\dot{g}^+ + \varepsilon^c. \tag{3.12}$$

Thus, unlike for the index-1 case, in the index-2 formulation an error in the algebraic constraints is amplified by $1/\tau$. This instability is observed in the numerical example in Section 4.

3.4. **Stable Discretization Schemes.** In this section we summarize the most common finite element schemes used in CFD. All mentioned schemes satisfy an inf-sup (also called Ladyzenskaja-Babuška-Brezzi) condition which is necessary to ensure stability of the pressure variable [11, Ch. VI.3]. Additional stable schemes are addressed in [16, Ch. II] as well as in [17, Ch. 3].

Using standard notation, we denote by $\mathcal{P}_k(\mathcal{T})$ the space of piecewise polynomials of degree $k$. The space of piecewise polynomials which are globally continuous is denoted by

$$\mathcal{S}_k(\mathcal{T}) := \mathcal{P}_k(\mathcal{T}) \cap H^1(\Omega).$$

With zero boundary conditions, we write $\mathcal{S}_{k,0}(\mathcal{T})$. For the pressure variable, we introduce the space

$$\mathcal{P}_0^0(\mathcal{T}) := \mathcal{P}_0(\mathcal{T})/\mathbb{R} = \mathcal{P}_0(\mathcal{T} \setminus \{T_0\})$$

for some triangle $T_0 \in \mathcal{T}$. Similarly, we define $\mathcal{S}_1^0(\mathcal{T}) := \mathcal{S}_1(\mathcal{T})/\mathbb{R}$. The discontinuous Crouzeix-Raviart finite element space [14] with zero boundary conditions is given by

$$\mathrm{CR}_0(\mathcal{T}) := \mathcal{P}_1(\mathcal{T}) \cap C(\{\mathrm{mid}(E) \mid E \in \mathcal{E}\}) \cap \{v \mid v(\mathrm{mid}(\mathcal{E}_{\mathrm{ext}})) = 0\}.$$

This space contains piecewise affine functions which are continuous at the midpoints of interior edges and vanish at the midpoints of boundary edges. Following [40], we introduce bubble functions on edges and triangles. An edge-bubble function is defined as scaled product of the two nodal hat functions corresponding to the two nodes of an edge. A triangle-bubble function is the scaled product of three nodal hat functions and thus, has local support on a single triangle. We denote this finite element space by $\mathcal{B}_3(\mathcal{T})$. Note that this is automatically a subspace of $H_0^1(\Omega)$.

It is well-known that the $[\mathcal{S}_{1,0}(\mathcal{T})]^2 - \mathcal{P}_0^0(\mathcal{T})$ scheme is not stable [11, Ex. VI.3.1]. An alternative low-order scheme was introduced in [14] and is given by

$$V_h = \big[\mathrm{CR}_0(\mathcal{T})\big]^2, \qquad Q_h = \mathcal{P}_0^0(\mathcal{T}). \tag{3.13}$$

In [23] yet another finite element space is introduced with less degrees of freedom by the mixture of continuous and discontinuous velocity components. This is given by

$$V_h = \mathcal{S}_{1,0}(\mathcal{T}) \times \mathrm{CR}_0(\mathcal{T}), \qquad Q_h = \mathcal{P}_0^0(\mathcal{T}).$$

An alternative approach is to enrich the discrete velocity space by bubble functions. Bernardi [8] uses edge-bubble functions multiplied by the outer normal vector of the corresponding edge. Thus, the fluxes through interior edges provide additional degrees of freedom. This ansatz is analysed in more detail in Section 3.5.2.

Using instead the triangle-bubble functions, we obtain the so-called MINI element [3, 11], defined by

$$V_h = \big[\mathcal{S}_{1,0}(\mathcal{T}) \oplus \mathcal{B}_3(\mathcal{T})\big]^2, \qquad Q_h = \mathcal{S}_1^0(\mathcal{T}).$$

Note that the ansatz for the pressure is continuous which yields a more natural model. Quite popular are approaches of Taylor-Hood type [35]. Therein,

the velocities are approximated by polynomials of one degree higher than the pressure. The Taylor-Hood element of lowest order is defined by

$$V_h = \left[\mathcal{S}_{2,0}(\mathcal{T})\right]^2, \qquad Q_h = \mathcal{S}_1^0(\mathcal{T}). \tag{3.14}$$

3.5. **Decompositions of $V_h$.** In this subsection, we derive decompositions of several finite elements schemes mentioned above such that Assumption 3.1 is satisfied. Since we do not deal with divergence-free elements, all resulting discretizations schemes for system (3.4) will be of nonconforming nature. Nonconforming finite element methods, for which the discrete space is no subspace of the continuous ansatz space, are analysed in [10, Ch. 10].

We show the construction of $V_{h,1}$ and $V_{h,2}$ by means of three examples.

3.5.1. *Discontinuous Velocity.* As first example, we consider the discontinuous Crouzeix-Raviart ansatz, introduced in (3.13). This ansatz is often used since it provides an efficient tool for CFD [7]. A proof of the inf-sup condition (3.5) is given in [7, 14].

Let $T_0 \in \mathcal{T}$ denote the triangle on which the pressure is fixed. The following algorithm defines a one-to-one mapping $\iota : \mathcal{T} \setminus \{T_0\} \to \mathcal{E}_{\text{int}}$ which allows to define the discrete space $V_{h,2}$.

*Algorithm* 3.1 (Mapping $\iota$). *Step 1:* Choose any $T \in \mathcal{T} \setminus \{T_0\}$ which shares an edge with $T_0$ and denote this edge by $E := T_0 \cap T \in \mathcal{E}_{\text{int}}$. Then, define $\iota(T) := E$ and $\mathcal{T}_R := \mathcal{T} \setminus \{T_0, T\}$. If $\mathcal{T}_R = \emptyset$, then stop.
*Step 2:* If $T$ from the previous step has an edge-neighbour in $\mathcal{T}_R$, then continue with Step 2a. Otherwise, go to Step 2b.

*Step 2a:* Select such a neighbouring triangle $S \in \mathcal{T}_R$ and set $E := T \cap S \in \mathcal{E}_{\text{int}}$. Furthermore, set $\iota(S) := E$ and $\mathcal{T}_R := \mathcal{T}_R \setminus \{S\}$. If $\mathcal{T}_R = \emptyset$, then stop. Otherwise, return to Step 2 with $T := S$.

*Step 2b:* Reset $T \in \mathcal{T} \setminus \mathcal{T}_R$ such that there exists an edge-neighbour in $\mathcal{T}_R$ and return to Step 2.

An illustration of the algorithm is shown in Figure 3.1.

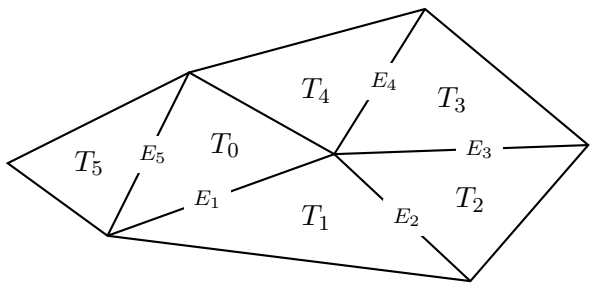


FIGURE 3.1. Illustration of Algorithm 3.1, $\iota(T_i) = E_i$ for $i = 1, \ldots, 5$. Step 2b of the algorithm is applied once to reset $T := T_0$.

*Remark* 3.3. Step 2b of Algorithm 3.1 is realizable since $\mathcal{T}_R \neq \emptyset$ and $\Omega$ is assumed to be connected with Lipschitz boundary. Furthermore, the algorithm terminates in finite time since the number of triangles is finite and Step 2a reduces the set of triangles $\mathcal{T}_R$ by one in at least every second iteration.

In the sequel, we will benefit of an order of the triangles $\mathcal{T} \setminus \{T_0\}$, given by their first appearence in Algorithm 3.1, namely $\{T_j\}_{j=1,\dots,|\mathcal{T}|-1}$.

Let $\phi_E$ denote the Crouzeix-Raviart basis function for an edge $E \in \operatorname{range}(\iota) \subset \mathcal{E}_{\text{int}}$, i.e., $\phi_E$ is piecewise linear with the value 1 at the midpoint of $E$ and 0 at the midpoint of any other edge. The corresponding triangle $T = \iota^{-1}(E)$ lies in the support of $\phi_E$ and thus, $\phi_E|_T$ cannot be constant. As a consequence, the divergence of either

$$\begin{bmatrix} \phi_E \\ 0 \end{bmatrix} \quad \text{or} \quad \begin{bmatrix} 0 \\ \phi_E \end{bmatrix}$$

has to be nonzero. Let $\Phi_E$ denote one of these basis funtions with $\operatorname{div}(\Phi_E|_T) \neq 0$. In the same manner, we select a basis function for every edge in the range of $\iota$ and obtain the ansatz space

$$V_{h,2} := \operatorname{span}\{\Phi_E \mid E \in \operatorname{range}(\iota)\}. \tag{3.15}$$

All remaining basis functions span the discrete space $V_{h,1}$. With the given decomposition of $V_h$, we obtain the following result.

**Lemma 3.2** (Decomposition for Crouzeix-Raviart)**.** *The discretization scheme $V_h - Q_h$ from* (3.13) *with the decomposition $V_h = V_{h,1} \oplus V_{h,2}$ defined in* (3.15) *satisfies Assumption 3.1.*

*Proof.* The matrix $B_2$ from Assumption 3.1 corresponds to the discrete space $V_{h,2}$ and is defined by

$$B_{2,ij} = \int_\Omega \chi_i \operatorname{div}(\Phi_j)\, \mathrm{d}x = \int_{T_i} \operatorname{div}(\Phi_j)\, \mathrm{d}x.$$

Therein, $\{\Phi_j\}$ denote the basis functions of $V_{h,2}$ and $\{\chi_i\}$ the basis functions of $Q_h$, i.e., $\chi_i = 1$ on the triangle $T_i$ and 0 elsewhere. Since $\operatorname{div}(\Phi_i) \neq 0$ on $T_i$ by construction, the diagonal entries of $B_2$ are nonzero. Furthermore, every column can only have two entries because of the support of edge-bubble functions. By the construction of Algorithm 3.1, the second entry can only be above the diagonal and thus, $B_2$ is upper triangular and nonsingular. □

3.5.2. *Continuous Velocity.* The second example applies a continuous approximation of the velocity but keeps the piecewise constants for the pressure as in (3.13). Since the $[\mathcal{S}_{1,0}(\mathcal{T})]^2 - \mathcal{P}_0^0(\mathcal{T})$ scheme is known to be unstable, the ansatz space $V_h$ is enriched by a special type of edge-bubble functions. As Bernardi [8, 16], we define for an interior edge $E \in \mathcal{E}_{\text{int}}$ the function

$$\Upsilon_E := \varphi_1 \varphi_2 \nu_E \in \mathcal{V}.$$

Therein, $\nu_E$ denotes the outer normal vector and $\varphi_1$, $\varphi_2$ the hat-functions corresponding the the vertices of the edge $E$. For an illustration of $\Upsilon_E$ see Figure 3.2. This yields as ansatz spaces for the velocity and the pressure,

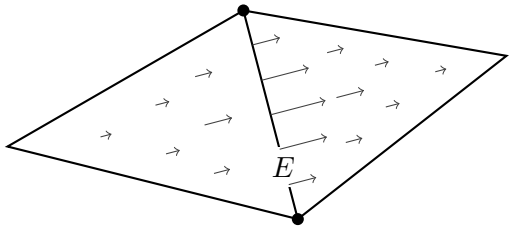


FIGURE 3.2. Illustration of the vector-valued function $\Upsilon_E = \varphi_1\varphi_2\nu_E$.

$$V_h = [\mathcal{S}_{1,0}(\mathcal{T})]^2 \oplus \{\Upsilon_E \mid E \in \mathcal{E}_{\text{int}}\}, \qquad Q_h = \mathcal{P}_0^0(\mathcal{T}). \tag{3.16}$$

The proof of the corresponding inf-sup condition can be found in [8]. In order to define the subspace $V_{h,2}$, we again use the mapping $\iota : \mathcal{T} \setminus \{T_0\} \to \mathcal{E}_{\text{int}}$ given by Algorithm 3.1. Therewith, we define

$$V_{h,2} = \{\Upsilon_E \mid E \in \text{range}(\iota)\} \tag{3.17}$$

and $V_{h,1}$ as the span of all remaining basis functions of $V_h$.

**Lemma 3.3** (Decomposition for Bernardi)**.** *The discretization scheme $V_h - Q_h$ from* (3.16) *with the given decomposition $V_h = V_{h,1} \oplus V_{h,2}$ defined in* (3.17) *satisfies Assumption 3.1.*

*Proof.* Note that the structure of $B_2$ is as in Lemma 3.2. Thus, it remains to show that the integral of $\operatorname{div} \Upsilon_E$ does not vanish. By definition of $\Upsilon_E$, it holds that

$$\operatorname{div} \Upsilon_E = \nabla(\varphi_1\varphi_2) \cdot \nu_E = \varphi_1 \nabla\varphi_2 \cdot \nu_E + \varphi_2 \nabla\varphi_1 \cdot \nu_E.$$

Hence, for a triangle $T$ with edge $E$,

$$\begin{aligned}\int_T \operatorname{div} \Upsilon_E \, \mathrm{d}x &= \nabla\varphi_2 \cdot \nu_E \int_T \varphi_1 \, \mathrm{d}x + \nabla\varphi_1 \cdot \nu_E \int_T \varphi_2 \, \mathrm{d}x \\ &= \frac{|T|}{3}\big(\nabla\varphi_2 + \nabla\varphi_1\big) \cdot \nu_E.\end{aligned}$$

Let $[x_i, y_i]^T$, $i = 1, 2$, denote the coordinates of the nodes corresponding to $\varphi_1$ and $\varphi_2$, respectively. W.l.o.g, we assume that the third node is located in $[0, 0]^T$. Then, the outer normal vector for $E$ is, up to a constant, given by $\nu_E = \begin{bmatrix} y_1 - y_2, & x_2 - x_1 \end{bmatrix}^T$. The hat-functions are defined by

$$\varphi_1(x, y) = \frac{1}{d}\big(y_2 x - x_2 y\big), \qquad \varphi_2(x, y) = \frac{1}{d}\big(-y_1 x + x_1 y\big)$$

with $d = x_1 y_2 - x_2 y_1 \neq 0$ since the triangle is part of a regular triangulation. Thus, we obtain

$$\big(\nabla\varphi_2 + \nabla\varphi_1\big) \cdot \nu_E = -\frac{1}{d}\big((x_1 - x_2)^2 + (y_1 - y_2)^2\big) \neq 0. \qquad \square$$

3.5.3. *Continuous Pressure.* This subsection is devoted to the decomposition of the popular Taylor-Hood element [35] in which the discretized pressure is continuous and the velocity is of higher order. The finite element spaces for this case are given in (3.14). The proof of the inf-sup condition is given in [39] or, using local arguments and macro elements, in [16, Ch. 2.4.2]. In the sequel, $v_p$ denotes the boundary node on which the pressure has no degree of freedom.

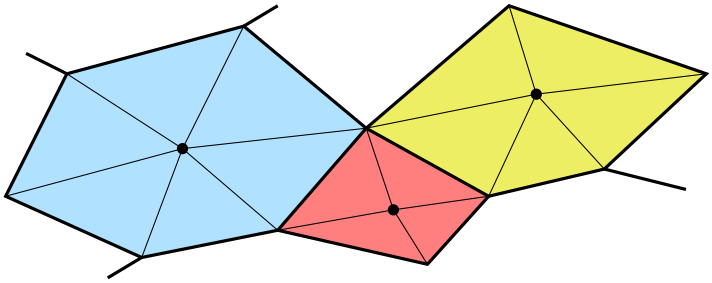

FIGURE 3.3. Sample of a triangulation $\mathcal{T}$ and its decomposition into three macro elements.

As in [16, Ch. II.4], we consider a triangulation $\mathcal{T}$ which can be decomposed into macro elements in the form of node patches (of interior nodes), see Figure 3.3. Thus, we assume that there exist macro elements $\{\Omega_r\}_{r=1,\dots,R}$, each with exactly one interior node, which form a partition of $\bar{\Omega}$. The triangulation of $\Omega_r$ is given by the restriction of $\mathcal{T}$ on $\bar{\Omega}_r$ and is denoted by $\mathcal{T}_r$. In addition, we assume that the macro elements are ordered such that $v_p$ is a node of $\mathcal{T}_1$ and that $\mathcal{T}_r$, $2 \le r \le R$, has a common node with at least one $\mathcal{T}_k$ for some $k \le r-1$. In order to decompose the finite element space $V_h$ such that Assumption 3.1 is fulfilled, we establish a one-to-one map $\jmath : \mathcal{N} \setminus \{v_p\} \to \mathcal{E}_{\text{int}}$. We define $\jmath$ by the following algorithm, which additionally introduces sets of nodes $\mathcal{I}_r \subset \mathcal{N}(\mathcal{T}_r)$.

*Algorithm* 3.2 (Mapping $\jmath$). Set $\mathcal{N}_R := \mathcal{N} \setminus \{v_p\}$. Iterate over macro elements, i.e., over $1 \le r \le R$:
*Step 1:* Consider the nodes $\mathcal{I}_r := \mathcal{N}_R \cap \mathcal{N}(\mathcal{T}_r) = \{v_0, \dots, v_{k(r)}\}$ where $v_0$ denotes the middle node, as shown in Figure 3.4.
*Step 2:* Define $E_j$ as the edge between $v_0$ and $v_j$ for $j = 1, \dots, k(r)$ and $E_0$ as any other edge of $\mathcal{E}(\mathcal{T}_r)$ which has $v_0$ as an endpoint.
*Step 3:* Set $\jmath(v_j) := E_j$ for $j = 0, \dots, k(r)$ and reset $\mathcal{N}_R := \mathcal{N}_R \setminus \mathcal{I}_r$. If $r \neq R$, return to Step 1 with $r := r+1$.

*Remark* 3.4. The order of the macro elements and the fact that $v_p \in \mathcal{N}(\mathcal{T}_1)$ guarantees that at least one node of $\mathcal{N}(\mathcal{T}_r)$ is not included in $\mathcal{N}_R \cap \mathcal{N}(\mathcal{T}_r)$. As a consequence, the second step of Algorithm 3.2 is always realizable.

It remains to define the subspaces $V_{h,1}$ and $V_{h,2}$. Similar to the previous decompositions, let $\Psi_E$ denote a function which vanishes in one component

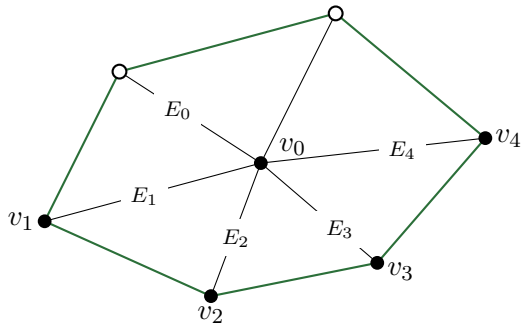


FIGURE 3.4. Single macro element $\Omega_r$ with an illustration of the map $\jmath : \mathcal{N} \setminus \{v_p\} \to \mathcal{E}_{\text{int}}$, $\jmath(v_k) = E_k$. Nodes of the type ○ are not part of $\mathcal{N}_r$ and thus, already covered by previous macro elements.

and equals the corresponding edge-bubble function $\psi_E$ in the other component. The precise order of the two components depends on the geometry, see the discussion for n-gons below. We then define

$$V_{h,2} := \operatorname{span}\{\Psi_E \mid E \in \operatorname{range}(\jmath)\}$$

and $V_{h,1}$ as the span of all remaining basis functions of $V_h$.

In the sequel, we denote by $B_{\mathcal{I}_r}$ the submatrix of $B$ which corresponds to the pressure nodes $\mathcal{I}_r$ (defined in Algorithm 3.2) and the edge-bubble functions of edges in $\jmath(\mathcal{I}_r)$.

**Lemma 3.4** (Localization of Taylor-Hood)**.** *Let $\mathcal{T}$ be a triangulation which can be decomposed into macro elements as illustrated in Figure 3.3. If all submatrices $B_{\mathcal{I}_r}$, $1 \le r \le R$, are invertible, then the discretization scheme* (3.14) *with the decomposition $V_h = V_{h,1} \oplus V_{h,2}$ satisfies Assumption 3.1.*

*Proof.* We show that the invertibility of the matrix $B_2$ from Assumption 3.1 follows from the invertibility of the local matrices. For this, the essential observation is that the ordering of macro elements together with Algorithm 3.2 leads to the block structure

$$B_2 = \begin{bmatrix} B_{\mathcal{I}_1} & * & * & * \\ & B_{\mathcal{I}_2} & * & * \\ & & \ddots & * \\ 0 & & & B_{\mathcal{I}_R} \end{bmatrix}. \qquad \square$$

*Remark* 3.5 ($n$-gons). On equlilateral $n$-gons, the invertibility of the submatrix $B_{\mathcal{I}_r}$ depends on a single parameter that we fix as the angle $\alpha$, enclosed by the edge $E_1$ and the $x$-axis, cf. Figure 3.5.

To see this, consider the column $B_{\mathcal{I}_r,\cdot j}$, associated with the edge-bubble function $\Psi_{E_j}$, $j = 0, 1, \dots, k(r)$. For illustration, assume that $\Psi_{E_j} = [\psi_{E_j}, 0]^T$.

Then, integration by parts yields for the entries

$$\begin{aligned} B_{\mathcal{I}_r,ij} = \int_{\Omega_r} \varphi_i \operatorname{div} \begin{bmatrix} \psi_{E_j} \\ 0 \end{bmatrix} \,\mathrm{d}x &= -\int_{T_{j-1}\cup T_j} \nabla\varphi_i \cdot \begin{bmatrix} \psi_{E_j} \\ 0 \end{bmatrix} \,\mathrm{d}x \\ &= -\frac{|T_{j-1}|}{3} \nabla\varphi_i|_{T_{j-1}} \cdot \begin{bmatrix} 1 \\ 0 \end{bmatrix} - \frac{|T_j|}{3} \nabla\varphi_i|_{T_j} \cdot \begin{bmatrix} 1 \\ 0 \end{bmatrix} \\ &= -\frac{1}{3} \int_{T_{j-1}\cup T_j} \operatorname{div} \begin{bmatrix} \varphi_i \\ 0 \end{bmatrix} \,\mathrm{d}x \\ &= -\frac{1}{3} \oint_{T_{j-1}\cup T_j} \begin{bmatrix} \varphi_i \\ 0 \end{bmatrix} \cdot \nu \,\mathrm{d}s. \end{aligned}$$

Thereby, we have made use of $\int_{T_j} \psi_{E_j} \,\mathrm{d}x = \frac{1}{3}|T_j|$, the local support of the basis functions, and the convention $T_{-1} = T_{n-1}$. On the contour of $T_{j-1}\cup T_j$, the outer normal $\nu$ is piecewise constant and given by $-n_{j-1}, m_{j-1}, m_j$, or $n_j$ from Figure 3.5, which depend solely on $\alpha$ and $n$.

From the integrals $\oint_{T_j} \varphi_i \,\mathrm{d}s$ one can factor out the radius, as the values of the integrals depend linearly on the length of the edges and since in an equilateral $n$-gon, the edges scale linearly with the radius and differ only by a factor depending on $n$. Since the radius can be factored out from every column of $B_{\mathcal{I}_r}$, invertibility depends only on $n$ and $\alpha$.

*Remark* 3.6. On an $n$-gon, let $\alpha_j$ denote the angle enclosed by the edge $E_j$ and the $x$-axis. Then, the decision rule

$$\Psi_{E_j} = \begin{bmatrix} 0 \\ \psi_{E_j} \end{bmatrix}, \text{ if } -\frac{\pi}{6} < \alpha_j + \ell\pi < \frac{\pi}{3}, \quad \Psi_{E_j} = \begin{bmatrix} \psi_{E_j} \\ 0 \end{bmatrix} \text{ otherwise}$$

for $\ell \in \mathbb{Z}$ and $j = 0, 1, \ldots, k(r)$ renders the via Algorithm 3.2 obtained matrix $B_{\mathcal{I}_r}$ invertible. This observation has been numerically proven correct for $3 \le n \le 9$ using the code available from the author's github account [20].

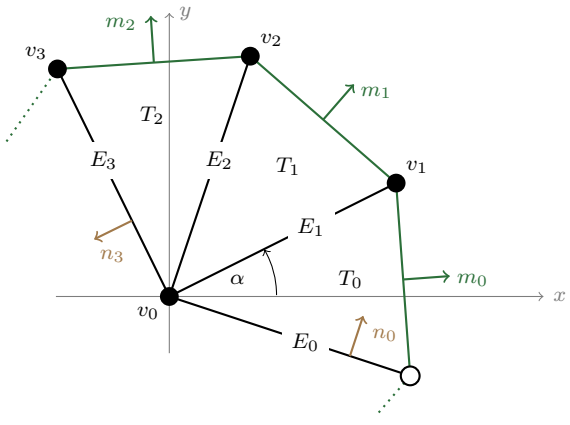


FIGURE 3.5. Notation for the equlilateral n-gon from Remark 3.5.

**Lemma 3.5** (Anisotropic scaling)**.** *Consider a patch $\Omega_r$ with nonsingular matrix $B_{\mathcal{I}_r}$. Then, $B_{\mathcal{I}_r}$ remains nonsingular under anisotropic scalings of $\Omega_r$, i.e., transformations of the form $S(x,y) = (ax, by)$ with $a, b > 0$.*

*Proof.* Let $\hat{\Omega}_r = S(\Omega_r)$ denote the transformed patch and $\hat{\varphi}_i$, $\hat{\Psi}_j$ the corresponding basis functions. Since $|\det \mathrm{D}\, S| = ab \neq 0$, the transformation formula gives for a transformed entry of $B_{\mathcal{I}_r}$,

$$\hat{B}_{\mathcal{I}_r,ij} = \int_{\hat{\Omega}_r} \hat{\varphi}_i \operatorname{div} \hat{\Psi}_j \, \mathrm{d}x = ab \int_{\Omega_r} \varphi_i \left( a \frac{\partial}{\partial x} \Psi_j + b \frac{\partial}{\partial y} \Psi_j \right) \mathrm{d}x.$$

Since $\Psi_j$ vanishes in one component, it holds $\hat{B}_{\mathcal{I}_r,ij} = c \cdot B_{\mathcal{I}_r,ij}$ with either $c = a^2 b$ or $c = ab^2$. In any case, this constant is the same for the entire column of $B_{\mathcal{I}_r}$ and thus, just a nonzero prefactor of the determinant. □

3.6. **Quadrilateral Meshes.** We close this section with a brief overview of stable finite element schemes on quadrilateral meshes and corresponding decompositions of the velocity space. Thus, the triangulation $\mathcal{T}$ is now supposed to be a partition of $\bar{\Omega}$ into convex quadrilaterals. For quadrilateral elements, one considers the space of piecewise polynomials of *partial* degree $k$ which are globally continuous, namely $\mathcal{Q}_k(\mathcal{T})$. With homogeneous boundary conditions we write $\mathcal{Q}_{k,0}(\mathcal{T})$.

As for the triangular case, there are finite element schemes of Taylor-Hood type, i.e., for $\ell \geq 2$,

$$V_h = \left[\mathcal{Q}_{\ell,0}(\mathcal{T})\right]^2, \qquad Q_h = \mathcal{Q}_\ell(\mathcal{T})/\mathbb{R}.$$

The stability result for this discretization can be found in [16, Ch. II.3.2]. It is well-known that the $[\mathcal{Q}_{1,0}(\mathcal{T})]^2 - \mathcal{P}_0(\mathcal{T})$ scheme is not stable. We refer to [9, Ch. III.7] for an illustration of the so-called checkerboard instability. Similar to the finite element scheme of Bernardi in Section 3.5.2, it is also possible to enrich the velocity space by the fluxes through interior edges. For a more precise characterization and the proof of stability, we refer to [16, Ch. II.3.1].

The last example we mention here is the analogon of the discontinuous approach of Crouzeix-Raviart. The element introduced by Rannacher and Turek [32] is given by

$$V_h = [\tilde{\mathcal{Q}}_{1,0}(\mathcal{T})]^2, \qquad Q_h = \mathcal{P}_0^0(\mathcal{T}). \tag{3.18}$$

Therein, $\tilde{\mathcal{Q}}_{1,0}$ denotes the nonconforming space which has one degree of freedom per interior edge. In contrast to the Crouzeix-Raviart element, functions in $V_h$ are not piecewise affine. Piecewise affine functions which are continuous in the midpoints of edges, were introduced by Park and Sheen [2, 30]. Unfortunately, there is no known stability result for this kind of element.

In a thorough analysis by Turek, the nonconforming element (3.18) was found superior over comparable conforming elements in terms of stability, accuracy, and efficiency [38, Ch. 3.1.1]. The higher stability and accuracy

of the nonconforming scheme is ascribed to the robustness of the inf-sup constant against mesh deformations.

A decomposition of $V_h$ from (3.18) into $V_{h,1}$ and $V_{h,2}$ in the sense of Assumption 3.1 works exactly as in Section 3.5.1.

## 4. Numerical Example

This section illustrates the benefits of the index-1 formulation (3.4) for numerical time integration by means of an example.

To isolate the high index effects, we consider a variant of (2.1) without the term $\frac{1}{\mathrm{Re}}\triangle u$ which introduces stiffness to the spatially discretized system and thus, step size restrictions for explicit schemes. As exact solution for the velocity field and the pressure in time and the two spatial coordinates, we set

$$\begin{aligned} u_1(t;x_1,x_2) &= 2\sin(8t)\cdot x_1^2(1-x_1)^2x_2(1-x_2)(2x_2-1),\\ u_2(t;x_1,x_2) &= 2\sin(8t)\cdot x_2^2(1-x_2)^2x_1(1-x_1)(1-2x_1),\\ p(t;x_1,x_2) &= \quad \sin(8t)\cdot x_1(1-x_1)x_2(1-x_2). \end{aligned}$$

The corresponding right-hand sides as well as the boundary and initial values are constructed accordingly. On the computational domain $(0,T)\times\Omega := (0,1)\times[0,1]^2$, this gives zero initial and zero Dirichlet boundary conditions.

The triangulation $\mathcal{T}_N$ of the spatial domain is characterized by the parameter $N$, meaning that the unit square is uniformly divided into $(N-1)^2$ squares that are clusters of four triangles each, see Figure 4.1. Besides, we choose as velocity and pressure state space the Taylor-Hood discretization (3.14).

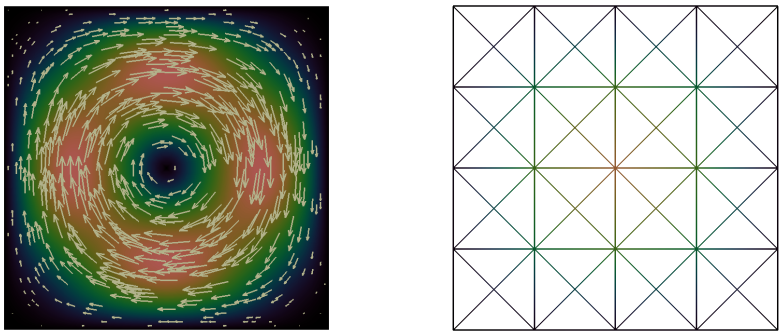

Figure 4.1. Illustration of the velocity field at $t=\frac{1}{4}$ and of the *criss-cross* triangulation for $N=5$.

This *criss-cross* triangulation and the Taylor-Hood elements enable the splitting $V_h = V_{h,1}\oplus V_{h,2}$ via Algorithm 3.2, cf. also Lemma 3.4 and Remark

3.6. Thus, the square matrix $B_2$ as defined in Assumption 3.1 is invertible and the index-1 formulation (3.4) is applicable. To investigate the time integration error, we discretize the time interval into $2^k + 1$, $k = 4, \dots, 10$, time instances and apply the semi-explicit Euler method. For the index-1 and index-2 formulations, this leads to the update formulas (3.9) and (3.8), respectively.

The resulting linear systems are solved iteratively up to a residual relative to the right-hand side ($rhs^c$) of the current iteration, i.e., until the norm of the residual is below $tol \cdot \min\{1, \|rhs^c\|\}$, for a given tolerance $tol$. For measuring the residuals we use the norms induced by inner product of the discrete $L^2$ spaces. For the index-2 system, this is the inner product with respect to the inverses of the mass matrices of the finite element bases. In the solution of the index-1 updates (3.9), where we used the block preconditioner with $[M_D^{-1}, B_2^{-1} M_{D,2} B_2^{-T}, B_2^{-1}]$ on the diagonal with $M_D$ denoting the diagonal of the mass matrix of the velocity space, this was approximated by the scalar product weighted by the mass matrices. Note, that the use of $B_2^{-1}$ is cheap, because of the blockdiagonal structure, cf. the proof of Lemma 3.4. It has turned out that for the numerical approximation of the index-2 formulation (3.8), it is beneficial to scale the differential equation by $\tau$.

By construction, the exact solution is known. The error of the numerical approximation is measured by taking the $L^2$-norm in space and evaluating the $L^2$-norm in time with the piecewise trapezoidal rule.

The numerical experiments show the improvements of the index 1 formulation for the pressure approximation, see Figure 4.2. As predicted by the theoretical considerations in (3.12), in the index-2 formulation, a numerical error in the algebraic constraints leads to a linear growth in the pressure error with decreasing time step sizes. A smaller residual in the continuity equation only postpones this instability. In the index-1 formulation, this systematic instability is not observed.

In theory, the pressure approximation error decreases linearly with $\tau$, [19, Ch. VII.4]. As in the investigations in [17], this is not observed here, because the error is dominated by the space discretization and $tol$.

The code used for the numerical investigation is available from the author's github account [20]. The finite element implementation uses *Fenics, Version 1.0.0,* [28], the linear systems are solved with *Krypy* [15].

## 5. Conclusion

We presented a new numerical approach to the unsteady NSE. Using analytical insight into the discrete spaces, we made the principles of the index reduction technique of minimal extension applicable to finite element schemes. In particular, the proposed variant preserves sparsity and maintains the physical meaning of the variables since only a permutation is applied. Unlike in penalization or in projection methods, our approach does not require time step restrictions or artificial boundary conditions for the pressure.

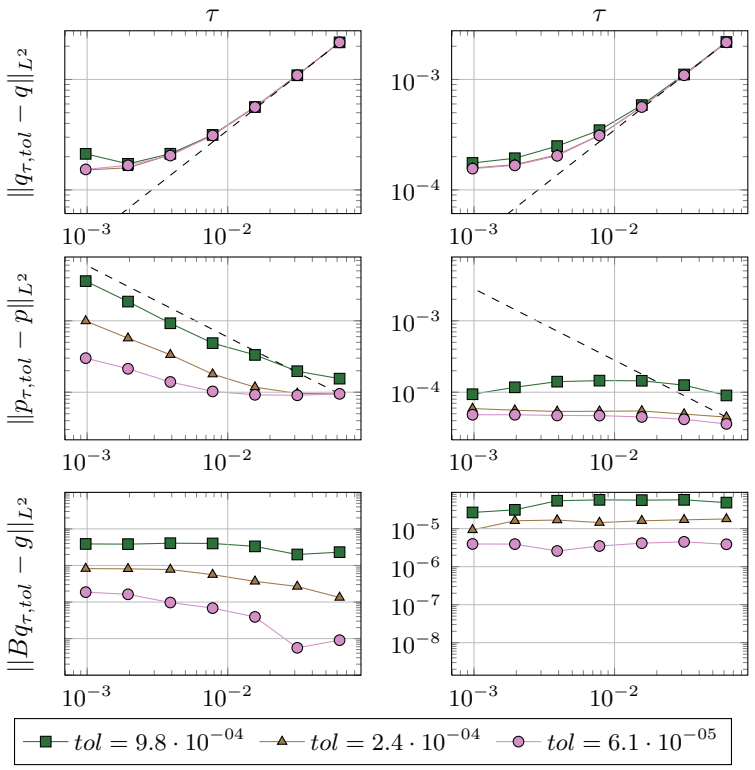


FIGURE 4.2. The evolution of the errors or residuals of the index-2 (left) and index-1 (right) formulation for varying time discretization parameter $\tau$ and *tol*. The space discretization is fixed with $N = 40$. The dashed lines is the linear fit.

The necessary splitting of the finite element space is operated for commonly used finite element discretizations as the Taylor-Hood and the Crouzeix-Raviart schemes. We have used the latter within within the numerical example that verifies the theoretical results.

We remark that the presented theory ensures consistency and stability of the half-explicit method. In view of efficiency, for practical computations, one need add a suitable preconditioner to solve system (3.9). In order to tackle stiff systems as the NSE for viscous fluids, one may consider a combination of our method with *IMEX* schemes that are implicit in the stiff linear part and explicit in the nonlinearity.

## REFERENCES

[1] R. Altmann. Index reduction for operator differential-algebraic equations in elastodynamics. *Z. Angew. Math. Mech. (ZAMM)*, pages (online), 10.1002/zamm.201200125, 2012.

[2] R. Altmann and C. Carstensen. $P_1$-nonconforming finite elements on triangulations into triangles and quadrilaterals. *SIAM J. Numer. Anal.*, 50(2):418–438, 2012.
[3] D. N. Arnold, F. Brezzi, and M. Fortin. A stable finite element for the Stokes equations. *Calcolo*, 21(4):337–344, 1984.
[4] M. Arnold. Half-explicit Runge-Kutta methods with explicit stages for differential-algebraic systems of index 2. *BIT*, 38(3):415–438, 1998.
[5] M. Arnold, K. Strehmel, and R. Weiner. Half-explicit Runge-Kutta methods for semi-explicit differential- algebraic equations of index 1. *Numer. Math.*, 64(4):409–431, 1993.
[6] U. M. Ascher, H. Chin, and S. Reich. Stabilization of DAEs and invariant manifolds. *Numer. Math.*, 67(2):131–149, 1994.
[7] R. Becker and S. Mao. Quasi-optimality of adaptive nonconforming finite element methods for the Stokes equations. *SIAM J. Numer. Anal.*, 49(3):970–991, 2011.
[8] C. Bernardi and G. Raugel. Analysis of some finite elements for the Stokes problem. *Math. Comp.*, 44(169):71–79, 1985.
[9] D. Braess. *Finite Elements - Theory, Fast Solvers, and Applications in Solid Mechanics*. Cambridge University Press, New York, third edition, 2007.
[10] S. C. Brenner and L. R. Scott. *The Mathematical Theory of Finite Element Methods*. Springer-Verlag, New York, third edition, 2008.
[11] F. Brezzi and M. Fortin. *Mixed and Hybrid Finite Element Methods*. Springer-Verlag, New York, 1991.
[12] S. L. Campbell and C. W. Gear. The index of general nonlinear DAEs. *Numer. Math.*, 72(2):173–196, 1995.
[13] P. G. Ciarlet. *The Finite Element Method for Elliptic Problems*. North-Holland, Amsterdam, 1978.
[14] M. Crouzeix and P.-A. Raviart. Conforming and nonconforming finite element methods for solving the stationary Stokes equations. I. *Rev. Franc. Automat. Inform. Rech Operat*, 7(R-3):33–75, 1973.
[15] A. Gaul. Krypy. Public Git Repository, Commit: 110a1fb756fb. Iterative Solvers for Linear Systems, `https://github.com/andrenarchy/krypy`.
[16] V. Girault and P.-A. Raviart. *Finite Element Methods for Navier-Stokes Equations*. Springer-Verlag, Berlin, 1986.
[17] P. M. Gresho and R. L. Sani. *Incompressible Flow and the Finite Element Method. Vol. 2: Isothermal Laminar Flow*. Wiley, Chichester, 2000.
[18] E. Hairer, C. Lubich, and M. Roche. *The Numerical Solution of Differential-Algebraic Systems by Runge-Kutta Methods*. Springer-Verlag, Berlin, 1989.
[19] E. Hairer and G. Wanner. *Solving Ordinary Differential Equations II: Stiff and Differential-Algebraic Problems*. Springer-Verlag, Berlin, second edition, 1996.

[20] J. Heiland. TayHoodMinExtForFlowEqns. Public Git Repository, Commit: 8eb641f21d. Solution of time-dependent 2D nonviscous flow with nonconforming minimal extension, `https://github.com/highlando/TayHoodMinExtForFlowEqns`.

[21] M. Heinkenschloss, D. C. Sorensen, and K. Sun. Balanced truncation model reduction for a class of descriptor systems with application to the Oseen equations. *SIAM J. Sci. Comput.*, 30(2):1038–1063, 2008.

[22] J. Heinrich and C. Vionnet. The penalty method for the navier-stokes equations. *Arch. Comput. Method E.*, 2:51–65, 1995.

[23] R. Kouhia and R. Stenberg. A linear nonconforming finite element method for nearly incompressible elasticity and Stokes flow. *Comput. Methods Appl. Mech. Engrg*, 124:195–212, 1995.

[24] P. Kunkel and V. Mehrmann. Index reduction for differential-algebraic equations by minimal extension. *Z. Angew. Math. Mech. (ZAMM)*, 84(9):579–597, 2004.

[25] P. Kunkel and V. Mehrmann. *Differential-Algebraic Equations: Analysis and Numerical Solution*. European Mathematical Society (EMS), Zürich, 2006.

[26] S. Le Borne and D. Cook II. Construction of a discrete divergence-free basis through orthogonal factorization in $\mathcal{H}$-arithmetic. *Computing*, 81(4):215–238, 2007.

[27] V. H. Linh and V. Mehrmann. Efficient integration of matrix-valued non-stiff DAEs by half-explicit methods. Preprint 2011–16, Technische Universität Berlin, Germany, 2011.

[28] A. Logg, K. Ølgaard, M. Rognes, and G. Wells. Ffc: the fenics form compiler. In *Automated Solution of Differential Equations by the Finite Element Method*, pages 227–238. Springer-Verlag, Berlin, 2012.

[29] G. Matthies and F. Schieweck. A multigrid method for incompressible flow problems using quasi divergence free functions. *SIAM J. Sci. Comput.*, 28(1):141–171, 2006.

[30] C. Park and D. Sheen. $P_1$-nonconforming quadrilateral finite element methods for second-order elliptic problems. *SIAM J. Numer. Anal.*, 41(2):624–640, 2003.

[31] A. Quarteroni and A. Valli. *Numerical Approximation of Partial Differential Equations*. Springer-Verlag, Berlin, 1994.

[32] R. Rannacher and S. Turek. Simple nonconforming quadrilateral Stokes element. *Numer. Meth. Part. D. E.*, 8(2):97–111, 1992.

[33] J. Shen. On error estimates of the penalty method for unsteady Navier-Stokes equations. *SIAM J. Numer. Anal.*, 32(2):386–403, 1995.

[34] L. Tartar. *An Introduction to Navier-Stokes Equation and Oceanography*. Springer-Verlag, Berlin, 2006.

[35] C. Taylor and P. Hood. A numerical solution of the Navier-Stokes equations using the finite element technique. *Internat. J. Comput. & Fluids*, 1(1):73–100, 1973.

[36] R. Temam. *Navier-Stokes Equations. Theory and Numerical Analysis.* North-Holland, Amsterdam, 1977.
[37] R. Temam. Remark on the pressure boundary condition for the projection method. *Theor. Comp. Fluid. Dyn.*, 3:181–184, 1991.
[38] S. Turek. *Efficient Solvers for Incompressible Flow Problems. An Algorithmic and Computational Approach.* Springer-Verlag, Berlin, 1999.
[39] R. Verfürth. Error estimates for a mixed finite element approximation of the Stokes equations. *RAIRO Anal. Numér.*, 18(2):175–182, 1984.
[40] R. Verfürth. *A Review of A Posteriori Error Estimation and Adaptive Mesh-Refinement Techniques.* Wiley-Teubner, Stuttgart, 1996.
[41] J. Weickert. *Applications of the Theory of Differential-Algebraic Equations to Partial Differential Equations of Fluid Dynamics.* PhD thesis, TU Chemnitz, Fakultät Mathematik, Chemnitz, Germany, 1997.

Institut für Mathematik MA4-5, Technische Universität Berlin, Strasse des 17. Juni 136, D–10623 Berlin, Germany

*E-mail address*: raltmann@math.tu-berlin.de, heiland@math.tu-berlin.de