CONTINUOUS, SEMI-DISCRETE, AND FULLY DISCRETISED NAVIER-STOKES EQUATIONS

R. ALTMANN[*] AND J. HEILAND[†]

ABSTRACT. The Navier-Stokes equations are commonly used to model and to simulate flow phenomena. We introduce the basic equations and discuss the standard methods for the spatial and temporal discretisation. We analyse the semi-discrete equations – a semi-explicit nonlinear DAE – in terms of the strangeness index and quantify the numerical difficulties in the fully discrete schemes, that are induced by the strangeness of the system. By analysing the Kronecker index of the difference-algebraic equations, that represent commonly and successfully used time stepping schemes for the Navier-Stokes equations, we show that those time-integration schemes factually remove the strangeness. The theoretical considerations are backed and illustrated by numerical examples.

Key words. Navier-Stokes equations, DAEs, strangeness index, difference-algebraic equations

AMS subject classifications. 65L80, 65M12, 35Q30

1. INTRODUCTION

The *Navier-Stokes equations* (NSE) are a system of nonlinear partial-differential equations that have been commonly used to model fluid flows for more than a century. The NSE are believed to describe all kinds of incompressible flows sufficiently well as long as the setup supports the hypothesis that the fluid is a continuum. Indeed, comparisons of numerical simulations with experiments show arbitrarily good agreement of the model with the observations over a long range from slowly moving flows in small geometries like a pipe up to highly turbulent flows over large spatial extensions like the flow around an airplane or even weather phenomena. Nevertheless, the mere equations and the extent of their applicability have not been fully deciphered by now and a substantial progress in this respect will be eligible for a Clay price[1].

Under the assumption of continuity of the observed quantities, the NSE can be derived from fundamental laws of physics; see [LL87] and [CM93]. These considerations are well backed for a macroscopic viewpoint, from which a fluid like water appears as a continuum. On a microscopic level, where discrete molecular structures define the states, the NSE cannot capture the physics right, as it is well-known, e.g., for capillary flows.

On the molecular level, fluids are better described by the *Boltzmann equations*, which model molecular interactions. This fact seems undisputed the more that the NSE can also be interpreted and derived through a limiting process of the Boltzmann equations in the sense of averaging the microscopic quantities for a macroscopic description [Sai09].

As a mathematical object the NSE have ever been subject to fundamental investigations and led to its own research field and its own subject definitions in the *MSC* classification scheme[2]. The research on the NSE has focused on the analysis of the equations and their numerical approximation. Early results on the existence of solutions are due to Leray [Ler33] (weak solutions) and Fujita&Kato [FK64] (smooth solutions). The first textbooks

---

Date: June 1, 2018.

[1]see http://www.claymath.org/millennium-problems/navier-stokes-equation

[2]see http://www.ams.org/mathscinet/msc/msc.html?t=35Q30&btn=Current and related

on the functional and on the numerical analysis were written by Ladyzhenskaya [Lad69] and Temam [Tem77], respectively.

For the numerical analysis of the spatial discretisation, one may distinguish two lines of development. The mathematical line focuses on Galerkin methods in the realm of variational formulations whereas the engineering orientated line has been advancing finite volume methods (FVM) as they appear well-suited for simulations. On the side of Galerkin methods, and in particular finite element methods (FEM), there have been many efforts in designing stable elements like the famous *Taylor-Hood* elements [TH73] as well as for general convergence results; see the textbooks [GR86, Lad69, Pir89] for the numerical analysis and [Tur99] for an application oriented overview. On the side of FVM that are the method of choice in most general purpose flow solvers, there have been general developments in view of discretising conservation laws [LeV92] and particular progress in view of stable approximation of fluid flow [FP02].

A numerical analysis of approximations to the time-dependent NSE with FEM semi-discretisations has been carried out by Heywood&Rannacher [HR90]; see also the textbook [GS00], that covers implementation issues. For time marching schemes for FVM formulations we refer to [FP02]. Strategies for the iterative solution of the arising linear systems can be found in [ESW05] (FEM) or [FP02] (FVM).

In this work we revisit the NSE for incompressible flows from a differential-algebraic equations (DAE) perspective. This includes the modelling where the incompressibility is treated as an algebraic constraint in an abstract space and the spatial semi-discretisation, that has to be handled with care to respect the incompressibility constraint and to lead to a well-posed classical DAE. It also includes the temporal discretisation, in which the DAE properties of the NSE become evident and (hopefully not) problematic. Such a pure DAE perspective has been taken on by Weickert [Wei97] who analysed finite difference approximations and certain time-stepping schemes for the NSE, by Emmrich&Mehrmann [EM13] who provided conditions and solution representations for linearised NSE in abstract spaces in line with linear time-invariant DAEs in finite-dimensional state-spaces. In [Hei14] the nonlinear NSE and its Galerkin approximations have been analysed in view of consistency of reformulations of semi-discrete DAE approximations with the infinite-dimensional model.

The paper is organized as follows. In Section 2 we derive the NSE from first principles and formulate the weak form as an *operator DAE*. The direct connection to DAEs is then made in Section 3, in which we report on several spatial discretisation schemes and that commonly used FEM schemes lead to systems of *strangeness index* one. In Section 4 we analyse the time approximation schemes in terms of the index of the resulting *difference-algebraic equations* (ΔAE). We show that the straight-forward temporal discretisation leads to a scheme of higher index than that of well-established time-stepping schemes for incompressible flows. We further confirm, that schemes with a ΔAE of lower index can also be obtained from a standard time-discretisation applied to a reformulation of the DAE with lower index. In the numerical examples in Section 5, we confirm the superiority of the lower index ΔAE approximations of the NSE over the straight-forward time-discretisation. As a benchmark for time-integration schemes for the considered class of nonlinear semi-explicit DAEs of strangeness index 1, we provide reference trajectories and the system coefficients and nonlinear inhomogeneities for direct realization in *Python*. We conclude this paper with summarizing remarks in Section 6.

2. Continuous model

2.1. **Derivation of the Navier-Stokes equations for incompressible flows.** The NSE provide a model of a flow as a continuum. The basic assumption for their derivation and, thus, the validity of the model is that the flow under consideration forms a continuous entity of flow particles in a spatial domain $\Omega \subset \mathbb{R}^3$ and a time interval $\mathcal{I}$ such that the functions

$$(2.1) \qquad \mathfrak{v} \colon \mathcal{I} \times \Omega \to \mathbb{R}^3, \quad \mathfrak{p} \colon \mathcal{I} \times \Omega \to \mathbb{R}, \quad \text{and} \quad \rho \colon \mathcal{I} \times \Omega \to \mathbb{R},$$

describing the velocity, the pressure, and the density as measured at the position $x \in \Omega$ at time $t \in \mathcal{I}$ are continuous functions. Here, continuous means that we can apply differential calculus in order to derive basic partial differential equations. Later, when we derive the weak formulation of the NSE, the needed continuity will be specified further.

As mentioned in [CM93, p. 2], these assumptions lead to a model that is believed to provide accurate descriptions of common macroscopic flow phenomena. In, e.g., setups of small geometric scales like in capillary flows or under vacuum-like conditions [KBN05], the discrete microscopic molecular structure of the fluid that constitutes the flow has to be taken into account.

The basic assumption of continuity of the matter allows for the consideration of a possibly infinitesimal small control volume $W \subset \Omega$, an open bounded domain in $\mathbb{R}^3$ that contains a given agglomerate of fluid particles in the considered flow. Continuity also implies that a fixed W is deformed and convected by the flow but always consists of the same fluid particles.

Under this continuity assumption, one can call on the *Reynolds Transport Theorem*, that relates the temporal change of an integral quantity over W to the convection velocity $\mathfrak{v}$.

Theorem 2.1 (see [Rey03], Eq.(16) and [CM93], p.10)*. Let $W \colon I \to \mathbb{R}^3$ describe a smoothly moving control volume and let $f \colon t \times W(t) \mapsto \mathbb{R}$ be a sufficiently smooth function, then*

$$(2.2) \qquad \frac{d}{dt} \int_W f \; \mathrm{d}V = \int_W \frac{\partial f}{\partial t} + \operatorname{div}(f\mathfrak{v}) \; \mathrm{d}V.$$

Incompressibility and mass conservation. The *Reynolds Transport Theorem* can be used to show that a flow is *incompressible*, which means that the volume of any agglomerate W is constant over time, if and only if the velocity field v is divergence free. In fact, if (2.2) applies, then one has that

$$(2.3) \qquad \frac{\mathrm{d}}{\mathrm{d}t} \int_W \; \mathrm{d}V = 0 \quad \text{if, and only if,} \quad \int_W \operatorname{div} \mathfrak{v} \; \mathrm{d}V = 0.$$

With a well-defined density function ρ, the mass of a control volume is defined as the integral over the (mass) density ρ and, with the assumption that in the flow there are neither mass sinks nor mass sources, an application of (2.2) gives

$$(2.4) \qquad \frac{\mathrm{d}}{\mathrm{d}t} \int_W \rho \; \mathrm{d}V = 0 \quad \text{if, and only if,} \quad \int_W \frac{\partial \rho}{\partial t} + \operatorname{div}(\rho\mathfrak{v}) \; \mathrm{d}V = 0.$$

Remark 2.2. The conservation of mass and the incompressibility of a flow are closely related and only equivalent in the case that the density ρ is constant in space and time. Flow models that assume incompressibility and varying density functions ρ are applied, e.g., in oceanography [Tar06].

Remark 2.3. If the volume of a fluid parcel changes over time, the flow is called *compressible*. In this case, the mass density ρ is modelled as an unknown function and related to the pressure p through constitutive or so-called *state equations*, like the *ideal gas law*; see [FKP16].

Balance of momentum. Under the continuum assumption, the momentum of a fluid agglomerate W can be expressed as the integral of the mass density times velocity and the temporal change equated with volume and surface forces on W:

$$(2.5) \qquad \frac{\mathrm{d}}{\mathrm{d}t} \int_W \rho \mathfrak{v} \ \mathrm{d}V = \int_W \rho g \ \mathrm{d}V + \int_{\partial W} \sigma n \ \mathrm{d}S,$$

where $g \colon \mathcal{I} \times \Omega \to \mathbb{R}^3$ is the density of a body force, and where $\sigma \colon \mathcal{I} \times \Omega \to \mathbb{R}^{3,3}$ is a tensor such that σn, where n is the normal field on the boundary ∂W of W, represents the density of the forces acting on the surface. Note that the expression in (2.5) is vector valued and that integration and differentiation is performed componentwise.

Applying the *Divergence Theorem* componentwise, one can write the term with the surface forces as a volume integral

$$(2.6) \qquad \int_{\partial W} \sigma n \ \mathrm{d}S = \int_W \mathrm{div} \, \sigma \ \mathrm{d}V,$$

with the divergence of a tensor defined accordingly.

So far, all assumptions have based on first principles. For the mathematical modelling of the tensor σ, however, ad hoc assumptions and heuristics are employed. First of all, it is assumed that σ can be written as

$$(2.7) \qquad \sigma = -\mathfrak{p}I + \tau$$

where $\mathfrak{p}$ – the pressure – is a smooth scalar function and $\tau \colon \mathcal{I} \times \Omega \to \mathbb{R}^{3,3}$ is the tensor of shear stresses. It is assumed that $\tau(t,x)$ is symmetric and invariant under rigid body rotation. Furthermore, τ is a linear function of the velocity gradient $\nabla \mathfrak{v} := \left[\frac{\partial \mathfrak{v}_j}{\partial x_i}\right]_{i,j=1,2,3}$. Under the additional assumption that the fluid under consideration is a *Newtonian fluid*, the tensor τ is defined by the velocity field and two further parameters (see [CM93, Ch. 1.3]) and commonly written as

$$(2.8) \qquad \tau = \mu \left[\nabla \mathfrak{v} + \nabla \mathfrak{v}^{\mathsf{T}} - \frac{2}{3}(\mathrm{div}\,\mathfrak{v})I \right] + \zeta(\mathrm{div}\,\mathfrak{v})I.$$

The parameter μ is the *(first coefficient of the) viscosity* and has been experimentally determined and tabulated for many gases and fluids. The parameter ζ is called the *bulk viscosity* and, in line with the *Stokes Hypothesis*, often set to zero.

As for the left hand side in (2.5), one proceeds as follows. Let $\mathfrak{v}_i$ denote the i-th component of $\mathfrak{v}$, $i = 1, 2, 3$. Then, an application of (2.2) gives that

$$(2.9) \quad \frac{\mathrm{d}}{\mathrm{d}t} \int_W \rho \mathfrak{v}_i \ \mathrm{d}V = \int_W \frac{\partial \rho \mathfrak{v}_i}{\partial t} + \mathrm{div}(\rho \mathfrak{v}_i \mathfrak{v}) \ \mathrm{d}V = \int_W \frac{\partial \rho \mathfrak{v}_i}{\partial t} + \rho \mathfrak{v}_i \, \mathrm{div}\, \mathfrak{v} + \nabla(\rho \mathfrak{v}_i) \cdot \mathfrak{v} \ \mathrm{d}V$$

where basic vector calculus has been applied. If one considers ∇ as the formal column vector of the three space derivatives, relation (2.9) for all components of v can be written in compact form as

$$(2.10) \qquad \frac{\mathrm{d}}{\mathrm{d}t} \int_W \rho \mathfrak{v} \ \mathrm{d}V = \int_W \frac{\partial \rho \mathfrak{v}}{\partial t} + \rho \mathfrak{v} \, \mathrm{div}\, \mathfrak{v} + (\mathfrak{v} \cdot \nabla)(\rho \mathfrak{v}) \ \mathrm{d}V.$$

The Navier-Stokes equations for incompressible flows. In the preceding derivations, integral quantities over a flow agglomerate W were considered. If the underlying *continuity assumption* includes that all relations hold on arbitrary (small) control volumes W, instead of equating the integrals, one can equate the integrands pointwise in space, which leads to partial-differential equations.

Thus, putting together all assumptions and derivations, the balance of momentum (2.5) for an isotropic Newtonian fluid under the *Stokes Hypothesis* defines the (NSE) as

$$(2.11) \qquad \frac{\partial \rho \mathfrak{v}}{\partial t} + \rho \mathfrak{v} \operatorname{div} \mathfrak{v} + (\mathfrak{v} \cdot \nabla)(\rho \mathfrak{v}) = \rho g - \nabla \mathfrak{p} + \operatorname{div}\big(\mu \big[\nabla \mathfrak{v} + \nabla \mathfrak{v}^{\mathsf{T}} - \frac{2}{3}(\operatorname{div} \mathfrak{v})I\big]\big).$$

In the case that the flow is incompressible with a constant density $\rho \equiv \rho^*$ and a constant viscosity $\mu \equiv \mu^*$, the combination of (2.4) and (2.11) results in the system

$$(2.12a) \qquad \rho^*\big(\frac{\partial \mathfrak{v}}{\partial t} + (\mathfrak{v} \cdot \nabla)\mathfrak{v}\big) + \nabla \mathfrak{p} - \mu^* \Delta \mathfrak{v} = \rho^* g,$$

$$(2.12b) \qquad \operatorname{div} \mathfrak{v} = 0.$$

Remark 2.4. Note that in the derivation, the pressure $\mathfrak{p}$ is not a variable but a function which is assumed to be known and to describe the normal forces on a fluid element, cf. (2.7) and [CM93, Ch. 1.1.ii]. Thus, the NSE (2.11) has only the velocity $\mathfrak{v}$ as an unknown and the divergence free formulation (2.12) can be seen as an abstract ODE for $\mathfrak{v}$ with an invariant [HLW06]. Nonetheless, since it turns out that the divergence constraint defines the function $\mathfrak{p}$, system (2.12) is commonly considered as an abstract differential-algebraic equation with $\mathfrak{v}$ and $\mathfrak{p}$ as unknowns.

For the numerical treatment and for similarity considerations, one relates all dependent and independent variables to a characteristic length L and characteristic velocity V via

$$(2.13) \qquad \mathfrak{v}' = \frac{\mathfrak{v}}{V}, \quad \mathfrak{p}' = \frac{\mathfrak{p}}{\rho^* U^2}, \quad x' = \frac{x}{L}, \quad \text{and} \quad t' = \frac{tV}{L}$$

With this and with $f' := \frac{U^2}{L} g$, equation (2.12) can be rewritten in dimensionless form

$$(2.14a) \qquad \frac{\partial}{\partial t'} \mathfrak{v}' + (\mathfrak{v}' \cdot \nabla)\mathfrak{v}' - \frac{1}{\operatorname{Re}} \Delta \mathfrak{v}' + \nabla \mathfrak{p}' = f',$$

$$(2.14b) \qquad \operatorname{div} \mathfrak{v}' = 0.$$

Thus, the system is completely parameterized by only one parameter $\operatorname{Re} := \frac{UL\rho^*}{\mu^*}$, the *Reynolds number*. In what follows, we will always consider the dimensionless NSE (2.14) but drop the dashes of the dimensionless variables.

Boundary conditions. For the considered domain Ω, let Γ denote the boundary in the abstract and in the physical sense. If the domain is bounded by nonpermeable walls, then the *no-slip* condition

$$(2.15) \qquad v = g \quad \text{on } \Gamma$$

applies, where g is the velocity of the wall. The no-slip conditions align well with experiments and macroscopic considerations; see [Lay08, Ch. 5.3] for references but also for examples where no-slip conditions seem insufficient to describe the flow at walls.

If the domain is not fully bounded by walls, typically because the computational domain needs to be bounded whereas the physical domain of the flow is unbounded, *artificial boundary* conditions are needed. If the direction of the flow is known, one may decompose these artificial boundaries into a part Γ_{i} for the incoming flow and Γ_{o} for the part where the flow leaves the domain. On Γ_{i} we can prescribe a velocity profile by

$$(2.16) \qquad \mathfrak{v} = \mathfrak{v}_i \quad \text{on } \Gamma_{\mathrm{i}},$$

whereas for Γ_{o}, there is no immediate physical insight into what should be the mathematical conditions. Accordingly, outflow boundary conditions are motivated as being *useful for the implementation of downstream boundary conditions* (see, e.g., [Glo03, Pir89]) and equipped

with the advice to put the outlet sufficiently far away from the region of interest; see [FP02, Ch.8.10.2]. The most common outflow boundary conditions are the *no stress* conditions:

$$\sigma n = 0 \quad \text{on } \Gamma_o,$$

where σ is the stress tensor as defined in its general form in (2.7), the *do-nothing* conditions:

$$(2.17) \qquad\qquad \frac{1}{\text{Re}} \frac{\partial \mathfrak{v}}{\partial n} - \mathfrak{p}n = 0 \quad \text{on } \Gamma_o,$$

that are formulated for the nondimensional NSE (2.14), or the *no gradient* conditions:

$$\frac{\partial \mathfrak{v}}{\partial n} = 0 \quad \text{and} \quad \frac{\partial \mathfrak{p}}{\partial n} = 0 \quad \text{on } \Gamma_o.$$

Remark 2.5. The *do-nothing* conditions have been extended; see [BM14], to the case of backflow, i.e., when some, possibly spurious, inflow occurs at the boundary Γ_o.

2.2. **Formulation as operator DAE.** In this subsection we provide yet another formulation of the dimensionless NSE (2.14), namely in the weak form as an operator DAE. This is a DAE in an abstract setting, where the solution is an element of a Sobolev space instead of a vector. We consider here homogeneous Dirichlet boundary conditions. More general boundary conditions with in- and outflow may modeled in a similar way, cf. [NR15]. We define the spaces

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

By $\mathcal{V}'$ we denote the dual space of $\mathcal{V}$. Note that the spaces $\mathcal{V}$, $\mathcal{H}$, $\mathcal{V}'$ form a Gelfand or evolution triple [Zei90, Ch. 23.4]. Furthermore, we define $\mathcal{W}(0,T)$ as the space of functions $u \in L^2(0,T;\mathcal{V})$, which contain a weak time derivative $\dot{u} \in L^2(0,T;\mathcal{V}')$. Well-known embedding results then imply $u \in C([0,T];\mathcal{H})$, cf. [Rou05, Lem. 7.3].

We now consider the weak formulation of (2.14) in operator form. This means that for 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 a pair $(\mathfrak{v},\mathfrak{p}) \in \mathcal{W}(0,T) \times L^2(0,T;\mathcal{Q})$ satisfying

$$(2.18a) \qquad\qquad \dot{\mathfrak{v}}(t) + \mathcal{K}(\mathfrak{v}(t)) - \mathcal{B}'\mathfrak{p}(t) = \mathcal{F}(t) \quad \text{in } \mathcal{V}',$$

$$(2.18b) \qquad\qquad \mathcal{B}\mathfrak{v}(t) \qquad\qquad = \mathcal{G}(t) \quad \text{in } \mathcal{Q}',$$

$$(2.18c) \qquad\qquad \mathfrak{v}(0) = a \qquad \text{in } \mathcal{H}$$

a.e. on $(0,T)$. The derivative of $\mathfrak{v}$ should be understood in the weak sense. The operators $\mathcal{K}\colon \mathcal{V} \to \mathcal{V}'$ and $\mathcal{B}\colon \mathcal{V} \to \mathcal{Q}'$ are defined via

$$(2.19) \qquad\qquad \langle \mathcal{K}(\mathfrak{v}), w \rangle = \int_\Omega (\mathfrak{v} \cdot \nabla)\mathfrak{v} \cdot w \, \mathrm{d}x + \frac{1}{\text{Re}} \int_\Omega \nabla \mathfrak{v} \cdot \nabla w \, \mathrm{d}x$$

and

$$(2.20) \qquad\qquad \langle \mathcal{B}\mathfrak{v}, q \rangle = \int_\Omega (\text{div } \mathfrak{v})q \, \mathrm{d}x = \langle \mathfrak{v}, \mathcal{B}'q \rangle,$$

respectively, for a given $\mathfrak{v} \in \mathcal{V}$ and for all test functions $w \in \mathcal{V}$ and $q \in \mathcal{Q}$. We emphasize that system (2.18) not only covers the NSE but also systems with time-dependent Dirichlet boundary conditions, since we have introduced an inhomogeneity $\mathcal{G}$, see [BH15b].

For results on the existence solutions to the weak formulation of the NSE, we refer to [Tar06]. For a compact summary that also considers nonzero $\mathcal{G}$ in (2.18b) see [AH15, AH18]. The differential-algebraic structure of (2.18) and the possible decoupling of differential and algebraic parts and variables has been discussed in [Hei14].

3. Semi-discrete equations

In this section, we consider the DAE, which results from a spatial discretisation of system (2.14) or (2.18):

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

$$(3.1b) \qquad\qquad Bv = g.$$

Here, v and p denote the finite-dimensional approximations of $\mathfrak{v}$ and $\mathfrak{p}$, respectively. In view of solvability, we assume additionally a consistent initial condition of the form $v(0) = v_0$, i.e., we demand $Bv_0 = g(0)$.

The apperance of B and B^T in (3.1) reflects the duality of the divergence and the gradient in the infinite-dimensional model that is preserved by mixed finite element discretizations of (2.18) (see [ESW05]) and finite difference discretizations of (2.14) (see [Wei97]). If finite volumes are used for the spatial discretization, typically, the coefficients in (3.1) are not explicitly assembled; see [FP02].

There are many reasons to include a right-hand side g in equation (3.1b) rather than a zero as it seems naturally for incompressible flows. Firstly, a nonzero g in the continuous continuity equations (2.14b) may appear also in generalizations of the NSE model to fluid-structure interactions (see, e.g., [Ray10]) or in optimal control setups; see, e.g., [Hin00]. Secondly, a nonzero g maybe a numerical artifact from the semi-discretisation of flows with nonzero Dirichlet boundary conditions as in the example we will provide below. Finally, in view of analysing the DAE, this inclusion of a nonzero g leads to a better understanding as one can track where the derivatives of the right-hand sides appear within the solution. This is of importance also in the case $g = 0$, since inexact solves lead to errors in the (3.1b) that act like a non-smooth inhomogeneity.

3.1. Spatial discretisation by finite elements.
We recall the basic principles of a discretisation of the incompressible Navier-Stokes equations by finite elements; see the text books [ESW05] for examples and [GR86] for the numerical analysis. Let V_h and Q_h be finite-dimensional subspaces of $\mathcal{V}$ and $\mathcal{Q}$, respectively, that are associated with a (shape) regular triangulation $\mathcal{T}$ of the polygonal Lipschitz domain Ω, cf. [Cia78]. Given a basis $\{\varphi_1, \ldots, \varphi_n\}$ of V_h, we can identify the finite-dimensional approximation of the velocity $\mathfrak{v}(t)$ by the coefficient vector $v(t) \in \mathbb{R}^n$. The discrete representative of the pressure $\mathfrak{p}(t)$ is denoted by $p(t) \in \mathbb{R}^m$ and corresponds to a basis $\{\psi_1, \ldots, \psi_m\}$ of Q_h.

With the basis functions of V_h we define the symmetric and positive definite mass matrix $M \in \mathbb{R}^{n,n}$ by

$$M := [m_{jk}] \in \mathbb{R}^{n \times n}, \qquad m_{jk} := \int_\Omega \varphi_j \cdot \varphi_k \, \mathrm{d}x.$$

The discretisation of the nonlinearity of the Navier-Stokes equation, i.e., the discretisation of the operator $\mathcal{K}$ in Section 2.2, is denoted by $K \colon \mathbb{R}^n \to \mathbb{R}^n$. Note that the given model also includes linearisations of the Navier-Stokes equation such as the unsteady Stokes or Oseen equation. In this case, K can be written as a $n \times n$ matrix. In view of the index analysis of system (3.1), however, this is not of importance; see Appendix A. We define for $v \in \mathbb{R}^n$ and its representative $\tilde{v} = \sum_{j=1}^n v_j \varphi_j \in V_h$,

$$(3.2) \qquad K_j(v) := \int_\Omega (\tilde{v} \cdot \nabla)\tilde{v} \cdot \varphi_j \, \mathrm{d}x + \frac{1}{\mathrm{Re}} \int_\Omega \nabla \tilde{v} \cdot \nabla \varphi_j \, \mathrm{d}x.$$

Finally, we define the matrix B, which corresponds to the divergence operator $\mathcal{B}$, i.e.,

$$B := [b_{ij}] \in \mathbb{R}^{m \times n}, \qquad b_{ij} := \langle \mathcal{B}\varphi_j, \psi_i \rangle = \int_\Omega \psi_i \operatorname{div} \varphi_j \, \mathrm{d}x.$$

By duality of the continuous operators, the discretisation of $\mathcal{B}'$ corresponds to the transpose B^T. This yields the saddle point structure of system (3.1).

Especially for the stable approximation of the pressure, it is necessary that the chosen finite element spaces are compatible [BF91, Ch. VI.3]. Let V_h and Q_h denote again finite-dimensional (sub)spaces of $\mathcal{V}$ and $\mathcal{Q}$, respectively, with

$$\dim V_h = n, \qquad \dim Q_h = m < n.$$

The spaces V_h and Q_h are compatible if they satisfy a so-called *inf-sup* or *Ladyzhenskaya-Babuška-Brezzi* condition [BF91, Ch. VI.3]. This means that there exists a constant $\beta > 0$, independent of the chosen mesh size, such that

$$\inf_{p_h \in Q_h} \sup_{v_h \in V_h} \frac{|\langle \operatorname{div} v_h, p_h \rangle|}{\|v_h\|_{\mathcal{V}} \|p_h\|_{\mathcal{Q}}} \geq \beta.$$

The inf-sup condition, with β independent of h, is a necessary condition for the convergence of the FEM; see, e.g., [Lad69] or [GR86]. For a fixed spatial discretisation, this condition implies that the matrix B resulting from the discretisation scheme V_h, Q_h is of full rank.

Remark 3.1. In the case of an internal flow, i.e., if there are no inflow or outflow boundaries, the pressure solution to (2.18) takes on values in $L^2(\Omega)/\mathbb{R}$ meaning that $\mathfrak{p}(t)$ is defined up to a constant. In a finite element discretisation that does not *fix* this constant explicitly, this leads to a rank deficit of B; see [ESW05, Ch. 5.3].

3.2. **Finite volumes and finite differences.** In this section we briefly touch the spatial discretisation of the incompressible NSE by the methods of *finite volumes* (FVM) and *finite differences* (FDM). If applied to (2.14) in a straight forward manner, both approaches lead to a DAE of type (3.1).

Finite volume approximations base on the integral formulation of the conservation laws as it reads for the momentum equation

$$(3.3) \qquad \int_W \frac{\partial \rho \mathfrak{v}}{\partial t} \, \mathrm{d}V + \int_{\partial W} \rho \mathfrak{v} \mathfrak{v} \cdot n \, \mathrm{d}S = -\int_{\partial W} \mathfrak{p} n \, \mathrm{d}S + \int_{\partial W} \tau \cdot n \, \mathrm{d}S + \int_W \rho g \, \mathrm{d}V,$$

cf. (2.5) and (2.7). Balances, like (3.3) in particular, hold for the whole domain of computation. For the discretisation, the domain of the flow is subdivided into small volumes often referred to as cells. The velocities and the pressure are assumed to be, say, constant over the cells, and their approximated values are determined by evaluating and equating the volume and surface integrals associated with every cell; see [FP02, Ch. 8.6].

Because of its flexibility in the discretisation, because of its variants that provide unconditional stability properties, and since typical models for the effective treatment of turbulence are formulated as conservation laws too, the FVM is the method of choice in most general purpose solvers. As for the discussion in the DAE context we note that, typically, the DAE (3.1) is never assembled but rather decoupled during the time discretisation; see [FP02, Ch. 7] and Section 4.3.2 below.

Approximations of the NSE via FDM come with the known flaws of finite difference approximations like high regularity requirements and confinement to regular grids. Nonetheless, FDM discretisations have been successfully used in flow discretisations and are probably still in use in certain specified, say, single-purpose codes. Also, FDM are often well suited for teaching the fundamentals of flow simulation of laminar and turbulent setups; see [GDN97].

3.3. **Index of the DAE.** As argued in Remark 3.1, the coefficient matrix B may be rank-deficient and, thus, the *differentiation index* may be not well-defined [Wei96]. However, the *strangeness index* [KM94, KM01, KM06] that applies to over- and underdetermined DAE systems and, thus, also can be determined in the case of a rank-deficient coefficient matrix B. A rough index analysis has been realized in [Wei97] with the result that under general and reasonable assumptions, system (3.1) has strangeness index 1. See also Appendix A, where the strangeness index has been determined in a rigorous way. This matches the, say, observations in [HW96, Ch. VII.1] that the system is of differentiation index 2 if the matrix B is of full rank.

In the sequel we always assume that the discretisation scheme is chosen in such a way that the matrix B is of full rank. Note that this can always be realized, choosing linearly independent finite element basis functions with respect to the space $\mathcal{Q}$, i.e., keeping in mind that constant functions are in the same equivalence class as the zero-function. If the ansatz functions form a partition of unity, then this means nothing else than eliminating one of these ansatz functions. Numerical schemes used in practice usually satisfy this condition as the discrete ansatz space for the pressure Q_h is chosen appropriately.

4. Fully discrete approximation schemes

The final step of a numerical approximation of the infinite-dimensional NSE (2.14) is the numerical time integration of the semi-discrete approximation (3.1). For this, one discretises the time interval $(0, T)$ via the grid

$$t_0 := 0 < t_1 < t_2 < \cdots < t_N = T$$

and computes a sequence $(v^k, p^k)_{k \in \mathbb{N}}$ of values that are supposed to approximate the solution of (3.1) at the discrete time instances, i.e., $v^k \approx v(t^k)$ and $p^k \approx p(t^k)$.

On the one hand side, this approximating sequence can be defined via a numerical time integration scheme applied to the DAE (3.1); see [KM06, Ch. 5] for a general introduction and [BH93, ASW93] for so-called *half-explicit* methods as we will consider them below. In this case, the resulting fully discrete scheme inherits the properties of the associated time-continuous DAE. In particular, if the DAE was strangeness-free, then suitable methods will deliver accurate and stable time-discrete approximations, whereas certain difficulties arise if the DAE is of higher index; cp. [KM06, Introduction to Ch. 6].

On the other hand side, the discrete sequence (v^k, p^k) may be defined as an approximation to (3.1) on a time grid without resorting to a time-discretization scheme. In this case, the resulting fully-discrete scheme cannot be assessed in terms of, say, the index of the associated time-continuous equations. Therefore, to provide a qualitative analysis of fully discrete schemes without resorting to the continuous-time system, we employ the Kronecker index for discrete-time systems.

4.1. **Index and causality of discrete systems.** For the analysis, we will restrict our considerations to a linear time-invariant setup. This is no restriction, since in the particular semi-linear semi-explicit form, the DAE structure is not affected by the nonlinearity. Further we assume that the time grid is equidistantly spaced and of size $\tau = t^{k+1} - t^k$. We will cast the fully-discrete schemes into the standard form

$$(4.1) \qquad\qquad \mathcal{E}x^{k+1} = \mathcal{A}x^k + h^k,$$

with coefficient matrices $\mathcal{E}$ and $\mathcal{A}$ and x^k containing all variables like $x^k = [v^k; p^k]$. We call such a fully discrete approximation scheme (4.1) a *difference-algebraic equation* (ΔAE).

Remark 4.1. This recast of the scheme, which will basically amount to a shift of the time indices for some variables, is needed to avoid ambiguities. In fact, for continuous time DAEs with delays, one can deliberately apply the operations of time-shift and differentiation to produce equivalent formulations of different indices [Ha15].

In what follows, we argue that the matrix pair $(\mathcal{E}, \mathcal{A})$ can be assumed to be regular (cf. [KM06, Def. 2.5]) such that it can be brought into a particular Kronecker form

$$(4.2) \qquad (\mathcal{E}, \mathcal{A}) \sim \left(\begin{bmatrix} I & 0 \\ 0 & N \end{bmatrix}, \begin{bmatrix} J & 0 \\ 0 & I \end{bmatrix} \right),$$

where J is a matrix in Jordan form and N is a nilpotent matrix, cf. [KM06, Thm. 2.7].

Definition 4.2. The index of nilpotency – that integer ν for which $N^{\nu} = 0$ while $N^{\nu-1} \neq 0$ – is called the *index of the matrix pair* $(\mathcal{E}, \mathcal{A})$; see [KM06, Def. 2.9], or the *Kronecker index* of the ΔAE (4.1), cf. [Rei06, p. 39].

Equivalently, we can classify the discrete schemes using the notion of *causality* that attributes systems with states depending only on the past or current inputs.

Definition 4.3 ([Dai89, Def. 8-1.1]). The sequence x^k, $k = 1, 2, \ldots$, defined through (4.1) is called *causal*, if x^k is determined completely by an initial condition x^0 and former (and the current) inputs $h^0, h^1, \ldots, h^k$.

Conversely, in a *noncausal* system, the current state x^k depends on future inputs like f^{k+1}. We will analyse time-stepping schemes for causality, which in the considered case is equivalent to being of Kronecker index 1 [Dai89, Thm. 8-1.1] and discuss how and why a noncausal system poses difficulties in the numerical approximation.

To introduce the procedure, to fix the notation, and to have a benchmark for further comparisons, we start with analysing an *half-explicit Euler* discretisation of a linearised version of (3.1), namely

$$(4.3a) \qquad \tfrac{1}{\tau} M v^{k+1} = (\tfrac{1}{\tau} M + A) v^k + B^T p^k + f^k,$$

$$(4.3b) \qquad B v^{k+1} = g^{k+1}.$$

This defines the difference equations for v^k and p^k, which approximate the velocity and pressure at the discrete time instances t^k, $k = 1, 2, \ldots$. Similarly, f^k and g^k stand for the approximations of f and g at the time instances.

The difference scheme is of the standard form (4.1) with $x^k = [v^k; p^k]$, $h^k = [f^k; g^k]$, a shift of the index in (4.3b), and

$$(4.4) \qquad (\mathcal{E}, \mathcal{A}) = \left(\begin{bmatrix} \tfrac{1}{\tau} M & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} \tfrac{1}{\tau} M + A & B^T \\ B & 0 \end{bmatrix} \right).$$

In Appendix B it is shown that under standard conditions and for τ sufficiently small, the pair $(\mathcal{E}, \mathcal{A})$ in (4.4) is regular and equivalent to

$$\left(\begin{bmatrix} I & 0 & 0 \\ 0 & 0 & 0 \\ 0 & I & 0 \end{bmatrix}, \begin{bmatrix} * & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & I \end{bmatrix} \right).$$

This means that the difference scheme based on an *half-explicit Euler* discretisation is of *Kronecker index* 2. In fact, a straight forward calculation reveals that in (4.3) the state p^k depends on g^{k+1}.

Remark 4.4. The index k for the pressure p in (4.3) is consistent with f^k, as can be directly derived from the case that $A = 0$ and $g = 0$. A fully implicit scheme, i.e., considering f^{k+1} and p^{k+1} in (4.3a), cannot be brought into the standard form (4.1), since a shift in the index, $k + 1 \leftarrow k$, would lead to the appearance of v^{k-1}.

4.2. **Inherent instabilities of ΔAEs of higher index.** In this section, we illustrate a mechanism that leads to a numerical instability and, thus, possibly to divergence of the approximation of a dynamical system through a time discretisation, i.e. a ΔAE, with a Kronecker index greater than 1. Consider a DAE in Kronecker form

$$N\dot{x} = x + g$$

with $N \neq 0$ and $N^2 = 0$ and it's time discrete approximation through an Euler scheme:

$$(4.5) \qquad \frac{1}{\tau}Nx^{k+1} = (\frac{1}{\tau}N + I)x^k + g^k.$$

The ΔAE (4.5) is of Kronecker index 2, according to Definition 4.2 and as it can be read off after a premultiplication by $(I - N)$, and it has the solution

$$x^k = -g^k - \frac{1}{\tau}(Ng^{k+1} - Ng^k),$$

as it follows from an adaption of the arguments in [KM06, Lem. 2.8] to the discrete case. From this solution representation one can conclude, that the solution to a ΔAE of higher index that discretises a DAE may depend on numerical differentiations and that any error in the computation may be amplified by the factor τ^{-1}. Note that for index-1 ΔAEs, where $N = 0$, this derivative is not present and that for even higher indices higher (numerical) derivatives will appear in the solution. Also note, that for systems that are not in Kronecker form such as (4.3), these derivations will be realized implicitly; see [AH15].

4.3. **Common time-stepping schemes as index-1 ΔAEs.** In this subsection, we discuss the different strategies, which are used in practice, to solve the spatially discretised NSE (3.1). In practical applications one typically uses schemes that decouple pressure and velocity computations. Although, as we have argued from a DAE perspective [AH15], this may lead to instabilities, the advantages that

- one has to solve two smaller systems rather than one large and
- one basically solves Poisson equations and convection-diffusion rather than saddle-point problems

seem to prevail despite of the active research on iterative solvers for saddle-point system. In particular, for the symmetric case of Stokes flow, theory provides efficient and well-understood preconditioning approaches. These approaches serve as a base for preconditioning techniques for coupled systems with highly non-normal coefficient matrices as they appear in the case of Navier-Stokes flow with higher Reynolds numbers; see, e.g., the numerical results in [ESW05, Ch. 8] or [HV16] for the scaling of the performance of iterative solvers with respect to the Reynolds number.

Another common feature of the schemes used in practice is their explicit approach to the momentum equation which avoids the repeated assembling of Jacobians. The error is then either controlled through a small time-step or through some fixed point iterations.

We consider the schemes *Projection* as it was described in [Gre90] and *SIMPLE* and *artificial compressibility*; see [FP02]. We will use a description and formulation general enough, to also accommodate numerous variants of the methods.

4.3.1. *Projection.* The principle of these methods is to solve for an intermediate velocity approximation that does not need to be divergence-free, and project it onto the divergence-free constraint in a second step. As far as the velocity approximation is concerned, *Projection* methods can be formulated both in infinite and finite dimensions. Since the first work by Chorin [Cho68], a number of variants have been developed mainly proposing different approaches to the approximation of the pressure tackling or circumventing the need of solving a *Poisson equation* for the pressure, which requires certain regularity assumptions (cf. [GS00, p. 642]). Another problem is the requirement of boundary conditions for the pressure update that do not have a physical motivation and may cause inaccuracies close to the boundary. Nonetheless, these schemes have been extensively studied and certain heuristics ensure satisfactory convergence behaviour; see, e.g, [Gre90, GS00].

As an example for a *Projection* scheme, we present the variant proposed in [Gre90]:

(1) Solve for intermediate velocity with the old pressure

$$(4.6) \qquad \frac{1}{\tau} M \tilde{v}^{k+1} = (\frac{1}{\tau} M + A) v^k + B^T p^k + f^k.$$

(2) Determine the new velocity v^{k+1} as the projection of the intermediate velocity onto $\ker B$ by solving

$$(4.7a) \qquad M v^{k+1} - B^T \phi^{k+1} = M \tilde{v}^{k+1},$$

$$(4.7b) \qquad B v^{k+1} = g^{k+1}.$$

(3) Update the pressure via

$$(4.8) \qquad p^{k+1} = p^k + \frac{2}{\tau} \phi^{k+1}.$$

Remark 4.5. Instead of solving the saddle-point problem (4.7) as a whole and in order to avoid the division of a numerically computed quantity by τ in (4.8), one can decouple the system. Then, one solves for $\tilde{\phi}^{k+1} := \frac{2}{\tau} \phi^{k+1}$ through

$$-B M^{-1} B^T \phi^{k+1} = \frac{2}{\tau} (B \tilde{v}^{k+1} - g^{k+1})$$

and obtains the updates via

$$v^{k+1} = \tilde{v}^{k+1} + \frac{\tau}{2} M^{-1} B^T \tilde{\phi}^{k+1}$$

and

$$p^{k+1} = p^k + \tilde{\phi}^{k+1}.$$

Remark 4.6. The formula for the update of the pressure is derived from the relation

$$\phi(t + \tau) = -\frac{\tau^2}{2} \dot{p}(t) + \mathcal{O}(\tau^3),$$

cf. [Gre90, p. 595], that holds under certain regularity assumptions. Note also that, if formulated for the space-continuous problem, this *Projection 2* algorithm requires a sophisticated treatment of the boundary conditions. The presented variant is a simplification for spatially discretised equations; see [GS00, Ch. 3.16.6c].

As a single system, this projection scheme defines a ΔAE in the form of (4.1) with $x^k := [\tilde{v}^k; \phi^k; v^k; p^k]$ and

$$(4.9) \qquad (\mathcal{E}, \mathcal{A}) = (\begin{bmatrix} \frac{1}{\tau} M & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ -M & -\frac{\tau}{2} B^T & M & 0 \\ 0 & -I & 0 & I \end{bmatrix}, \begin{bmatrix} 0 & \frac{1}{\tau} M + A & 0 & B^T \\ -\frac{2}{\tau} B & -B M^{-1} B^T & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & I \end{bmatrix}),$$

which is a matrix pair of Kronecker index 1; see Appendix B.2.

4.3.2. *SIMPLE scheme – implicit pressure correction.* The SIMPLE scheme and its variants are based on the decomposition

$$(4.10) \qquad v^{k+1} = \tilde{v}^{k+1} + v_\Delta^{k+1} \quad \text{and} \quad p^{k+1} = p^k + p_\Delta^{k+1},$$

where $\tilde{v}^{k+1}$ is the tentative velocity computed by means of the old pressure. We present the basic variant, in which the velocity correction v_Δ^{k+1} is discarded when solving for p_Δ^{k+1}:

(1) Solve for the intermediate velocity $\tilde{v}^{k+1}$ with the old pressure as in (4.6).
(2) Compute p_Δ^{k+1} through

$$(4.11) \qquad B(\tfrac{1}{\tau}M + A)^{-1}B^T p_\Delta^{k+1} = -B\tilde{v}^{k+1} + g^{k+1}$$

as the correction to p^k such that the
(3) updates of the velocity and pressure defined through

$$(4.12a) \qquad v^{k+1} = \tilde{v}^{k+1} + (\tfrac{1}{\tau}M + A)^{-1}B^T p_\Delta^{k+1},$$

$$(4.12b) \qquad p^{k+1} = p^k + p_\Delta^{k+1}.$$

jointly fulfill the time discrete momentum and the continuity equation.

Remark 4.7. Step (2) of the presented SIMPLE algorithm computes p_Δ^{k+1} under the temporary assumption that $v_\Delta^{k+1} = 0$. This is hardly justified and probably a reason for slow convergence, when applied in a fixed-point iteration within fully implicit schemes. Certain variants of the SIMPLE scheme try to approximate v_Δ^{k+1} at this step, e.g., through interpolation; see [FP02, Ch. 7.3.4].

As a single system, the SIMPLE scheme defines a ΔAE in the form of (4.1) with $x^k := [\tilde{v}^k; p_\Delta^k; v^k; p^k]$ and
(4.13)

$$(\mathcal{E}, \mathcal{A}) = (\begin{bmatrix} \tfrac{1}{\tau}M & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ I & (\tfrac{1}{\tau}M + A)^{-1}B^T & -I & 0 \\ 0 & -I & 0 & I \end{bmatrix}, \begin{bmatrix} 0 & \tfrac{1}{\tau}M + A & 0 & B^T \\ -B & -B(\tfrac{1}{\tau}M + A)^{-1}B^T & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & I \end{bmatrix}),$$

which is a matrix pair of Kronecker index 1; see Appendix B.3.

4.3.3. *Artificial compressibility.* In this class of methods, the divergence-free constraint (2.14b) is relaxed by adding a scaled time-derivative of the pressure, i.e.,

$$(4.14) \qquad \frac{1}{\beta}\frac{\partial \mathfrak{p}}{\partial t} + \operatorname{div} \mathfrak{v} = 0.$$

The parameter is assumed to satisfy $\beta \gg 1$. The corresponding spatial semi-discretisation (3.1) then reads

$$(4.15a) \qquad M\dot{v} + K(v) - B^T p = f,$$

$$(4.15b) \qquad M_p\dot{p} + Bv = g$$

with M_p denoting the mass matrix of the pressure approximation.

In theory, system (4.15) is an ODE and could be solved by standard time-stepping schemes. In practice, however, the common solution approaches to System (4.15) decouple pressure and velocity computations through introducing auxiliary quantities and, thus, a DAE structure. A possible method, that defines a pressure update similar to the SIMPLE scheme (4.12), can be interpreted and implemented as follows, cf. [FP02, Ch. 7.4.3]. As

in the SIMPLE approach a first velocity approximation $\tilde{v}^{k+1}$ is computed via (4.6) on the base of the old pressure value p^k.

Next, one substracts the contribution of the old pressure,

$$(4.16) \qquad \bar{v}^{k+1} := \tilde{v}^{k+1} - (\frac{1}{\tau}M + A)^{-1}B^T p^k$$

and defines the new velocity v^{k+1} through the linear expansion of $\bar{v}^{k+1}$ in terms of the pressure gradient, i.e.,

$$(4.17) \qquad v^{k+1} = \bar{v}^{k+1} + \frac{\partial(\bar{v}^{k+1})}{\partial(B^T p)}[B^T p^{k+1} - B^T p^k] = \bar{v}^{k+1} + (\frac{1}{\tau}M + A)^{-1}B^T p_\Delta^{k+1}.$$

Note that one uses (4.16) to determine the Jacobian $\frac{\partial(\bar{v}^{k+1})}{\partial(B^T p)}$ and that we have defined $p_\Delta^{k+1} = p^{k+1} - p^k$. Expression (4.17) is then inserted in a, e.g., *implicit Euler* discretisation of (4.15b), which gives an equation for p_Δ^{k+1}. Accordingly, the fully discrete scheme using artificial compressibility reads

$$(4.18a) \qquad (\tfrac{1}{\tau}M + A)\tilde{v}^{k+1} = \tfrac{1}{\tau}Mv^k + B^T p^k + f^k,$$

$$(4.18b) \qquad \bar{v}^{k+1} = \tilde{v}^{k+1} - (\tfrac{1}{\tau}M + A)^{-1}B^T p^k,$$

$$(4.18c) \qquad \frac{1}{\beta\tau}p_\Delta^{k+1} + B\bar{v}^{k+1} + B(\tfrac{1}{\tau}M + A)^{-1}B^T p_\Delta^{k+1} = g^{k+1},$$

$$(4.18d) \qquad v^{k+1} = \bar{v}^{k+1} - (\tfrac{1}{\tau}M + A)^{-1}B^T p_\Delta^{k+1},$$

$$(4.18e) \qquad p^{k+1} = p^k + p_\Delta^{k+1}.$$

The corresponding ΔAE equals the ΔAE of the SIMPLE scheme up to a slight modification of the equation for the pressure update such that we can use the arguments laid out in Appendix B.3 to conclude that it is of Kronecker index 1.

Remark 4.8. Note that in practice there may be nonlinear solves to determine, e.g., $\tilde{v}^{k+1}$ such that, e.g., the correction $\bar{v}^{k+1}$ is possibly different from $\tilde{v}^{k+1}$ computed with $p^k = 0$.

4.4. **Time-stepping schemes resulting from index reduction.** Besides the presented schemes in Section 4.3, one may also apply an index reduction to system (3.1) and then discretise in time. This then also leads to matrix pairs $(\mathcal{E}, \mathcal{A})$ of Kronecker index 1.

4.4.1. *Penalty methods.* Similar to Section 4.3.3, we may reduce the index of the DAE (3.1) by relaxing the divergence-free constraint or, in other words, add a penalty term [She95]. With the penalty parameter $\beta \gg 1$ we replace the incompressibility condition by

$$\mathfrak{p} = -\beta \operatorname{div} \mathfrak{v}.$$

In the semi-discrete case this corresponds to the constraint equation $M_p p + Bv = g$. It can be shown that this then leads to a DAE of differentiation index 1. However, this approach changes the solution of the system. In order to keep this difference of reasonable size, β should be choosen relatively large. On the other hand, the condition number of the involved matrices increase with β and thus, lead to numerical difficulties. The difficulty of a reasonable choice of β is one drawback of this approach. A second disadvantage is that small velocities of order β^{-1} or less cannot be resolved [HV95]. Further modifications of the penalty method, which are also applicable for slightly compressible fluids (Newtonian fluids)), are discussed, e.g., in [HV95, Sect. 5].

4.4.2. *Derivative of the constraint.* Another very simple possibility to reduce the index of the given DAE is to replace the constraint by its derivative, the so-called *hidden constraint*. Instead of (3.1) we then consider the system

$$(4.19a) \qquad\qquad M\dot{\tilde{v}} + K(\tilde{v}) - B^T\tilde{p} = f,$$

$$(4.19b) \qquad\qquad B\dot{\tilde{v}} = \dot{g}.$$

The initial condition remains unchanged, i.e., $\tilde{v}(0) = v_0$. It is well-known that this system has differentiation index 1. Nevertheless, numerical simulations, which rely on this formulation, show a linear drift from the solution manifold given by $Bv = g$. This can be seen as follows.

Although the two systems are equivalent, numerical errors are integrated over time and thus amplified. Solving the constraints only up to a small error, i.e.,

$$Bv(t) = g(t) + \varepsilon, \qquad B\dot{v}(t) = \dot{g}(t) + \eta$$

for small and constant ε and η, we calculate for $\tilde{v}$ that

$$B\tilde{v}(t) = Bv_0 + \int_0^t B\dot{\tilde{v}}(s)\,\mathrm{d}s = Bv_0 + \int_0^t \left(\dot{g}(s) + \eta\right)\mathrm{d}s = g(t) + t\,\eta.$$

Note that the last step holds because of the assumed consistency condition $Bv_0 = g(0)$. This shows that a constant error in the constraint of $\tilde{v}$ leads to an error which grows linearly in time. Because of this, the method of replacing the constraint by its derivative is – although it is of lower index – not advisable.

4.4.3. *Minimal extension.* Finally, we present the index reduction technique of *minimal extension* [KM06, Ch. 6.4]. A general framework for an index reduction based on derivative arrays is given in [KM06, Ch. 6]; see also [Cam87]. Because of the special saddle point structure of system (3.1), in which the constraint is explicitly given, this procedure can be simplified by using so-called *dummy variables*, cf. [MS93, KM04].

Since we assume that B is of full rank, there exists an invertible matrix $Q \in \mathbb{R}^{n,n}$ such that BQ has the block structure $BQ = [B_1,\ B_2]$ with an invertible matrix $B_2 \in \mathbb{R}^{m,m}$. Note that the choice of Q is not unique. We then use this transformation to partition the variable v, namely

$$\begin{bmatrix} v_1 \\ v_2 \end{bmatrix} := Q^{-1}v,$$

with according dimensions $v_1 \in \mathbb{R}^{n-m}$ and $v_2 \in \mathbb{R}^m$. With this, the constraint and its derivative may be written as

$$B_2 v_2 = g - B_1 v_1, \qquad B_2\dot{v}_2 = \dot{g} - B_1\dot{v}_1.$$

Together with the differential equation (3.1a) this yields an overdetermined system. Instead of an (expensive) search for projectors [KM06, Ch. 6.2] or selectors [Ste06], which would bring the system back to its original size, we introduce a dummy variable $w_2 := \dot{v}_2$ to get rid of the redundancy. Note, however, that this leads to a slightly bigger system. More precisely, we extend the system dimensions from $n + m$ to $n + 2m$, which is still moderate, since we usually have $m \ll n$. The extended system is again square and has the form

$$MQ\begin{bmatrix} \dot{v}_1 \\ w_2 \end{bmatrix} + K(v_1, v_2) - B^T p = f,$$

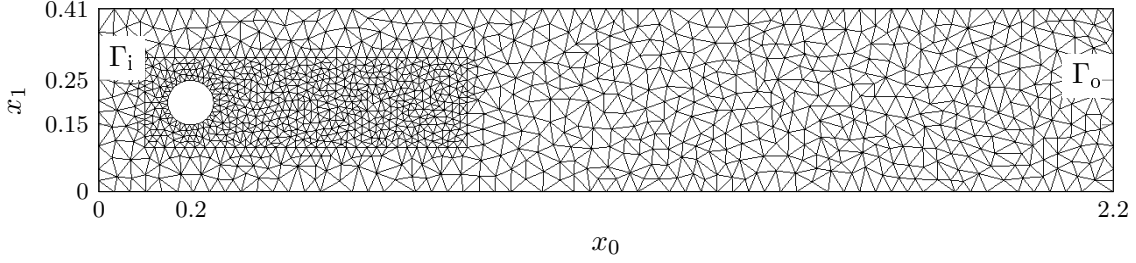$$B_2 v_2 = g - B_1 v_1,$$

$$B_2 w_2 = \dot{g} - B_1\dot{v}_1.$$

FIGURE 5.1. Illustration of the geometrical setup including the domains of distributed control and observation and of the velocity magnitude for the cylinder wake. Figure taken from [BBH17].

This DAE is of differentiation index 1 and has the same solution set as the original system, cf. [AH15].

The drawback of this method is the need of a transformation matrix Q. With a suitable reordering of the variables, however, we can choose Q to be the identity matrix. In this case, the needed variable transformation is just a permutation and thus, all variables keep their physical meaning. For some specific finite element schemes an algorithm to find such a permutation (which leads to an invertible B_2 block) is given in [AH15]. This paper also considers the half-explicit Euler scheme applied to the minimally extended system, which leads to a pair $(\mathcal{E}, \mathcal{A})$ of Kronecker index 1.

5. NUMERICAL EXPERIMENTS

To illustrate the performance and particular issues of the time-stepping schemes for the NSE, we consider the numerical simulation of the flow passing a cylinder in two space dimensions. This problem, also known as *cylinder wake*, is a popular flow benchmark problem and a test field for flow control [Wil96, NAM+03, BH15a].

We consider the incompressible NSE (2.14) on the domain as illustrated in Figure 5.1 with boundary Γ and boundary conditions as follows. At the inflow Γ_i, we prescribe a parabolic velocity profile through the function

$$g(s) = 4\big(1 - \frac{s}{0.41}\big)\frac{s}{0.41}.$$

At the outflow Γ_o, we impose *do-nothing* conditions, cf. (2.17). At the upper and the lower wall of the channel and at the cylinder periphery, we employ *no-slip*, i.e., zero Dirichlet conditions. Thus, we set

$$\gamma(v,p)\colon \begin{cases} v = [g(x_1), 0]^T & \text{on } \Gamma_i, \\ p\vec{n} - \frac{1}{\text{Re}}\frac{\partial v}{\partial \vec{n}} = [0,0]^T & \text{on } \Gamma_o, \\ v = [0,0]^T & \text{elsewhere on the boundary.} \end{cases}$$

We set Re = 60 and consider a $\mathcal{P}_2$–$\mathcal{P}_1$ (Taylor-Hood) finite element discretisation of (2.14) on the grid depicted in Figure 5.1 with 9356 degrees of freedom in the velocity and 1289 degrees of freedom in the pressure approximation. The result of the semi-discretisation is a DAE of the form (3.1), which we write as

(5.1a) $$M\dot{v} + Av + N(v) - B^T p = f,$$
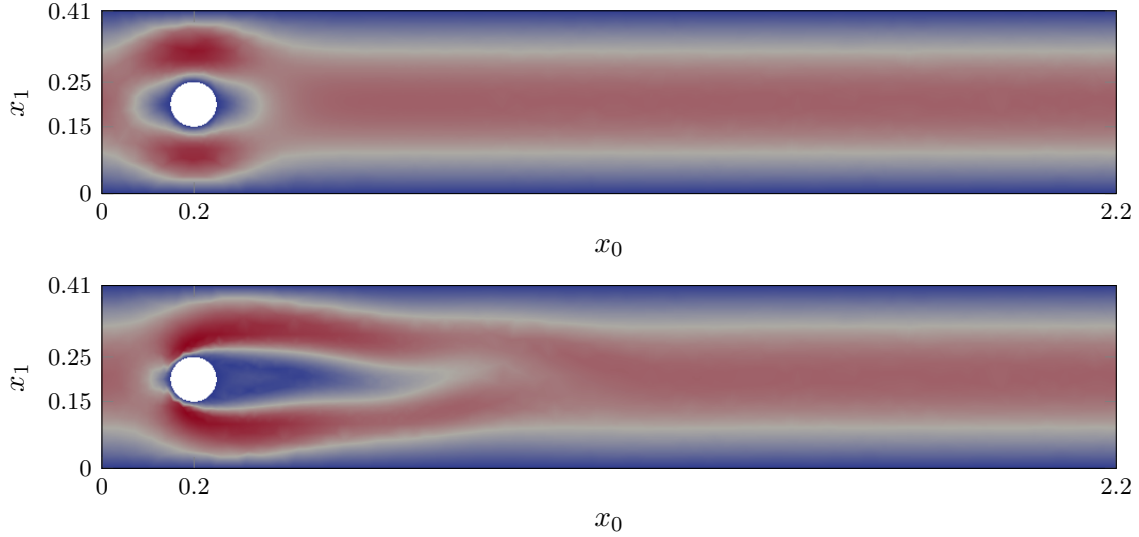
(5.1b) $$Bv = g$$

FIGURE 5.2. Snapshots of the velocity magnitude computed with SIMPLE with $\tau = 1/1024$ and exact solves taken at $t = 0$ (top) and $t = 1$ (bottom).

on the time interval $(0, 1]$. Here, A is the Laplacian or the linear part of K as defined in (3.2) and N denotes the convection part. The right-hand sides f and g account for the (static) boundary conditions. As initial value we take the corresponding steady-state *Stokes* solution v_S, which is part of the solution to

$$(5.2) \qquad \begin{bmatrix} A & -B^T \\ B & 0 \end{bmatrix} \begin{bmatrix} v_\mathrm{S} \\ p_\mathrm{S} \end{bmatrix} = \begin{bmatrix} f \\ g \end{bmatrix}.$$

For the schemes that need an initial value for the pressure, we provide it as p_NS solving (5.1) at $t = 0$ with $v(0) = v_\mathrm{S}$ and $\dot{v}(0) = 0$. Note that this gives a consistent initial pressure, since g is constant, and thus, $B\dot{v} = 0$.

We consider the time-discretisation of (5.1) by means of the half-explicit Euler scheme and compare it to the time-discretisation via the SIMPLE scheme as described in Section 4.3.2. Both schemes treat the nonlinearity explicitly. Comparing the computed approximations to a reference, obtained by the time-discretisation via the implicit trapezoidal rule on a fine grid, we show that both schemes are convergent of order 1 as long as the resulting linear systems are solved with sufficient precision. For inexact solves, we show that the pressure approximation in the half-explicit Euler scheme diverges unlike for the SIMPLE approximation. For snapshots of the approximate velocity solution, see Figure 5.2.

The finite element implementation uses *FEniCS, Version 2017.2* [LORW12]. For the iterative solutions of the linear system, we employ *Krypy* [Gau17]. The code used for the numerical investigations is freely available for reproducing the reported results and as a benchmark for further developments in the time integration of semi-explicit DAEs of strangeness index 1; see Figure 5.3 for a stable link to the online repository.
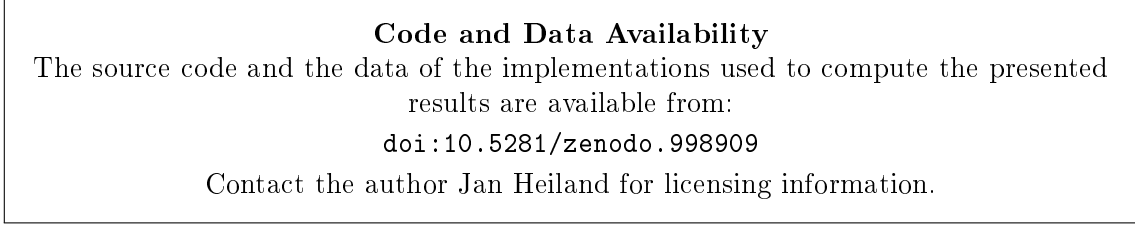
FIGURE 5.3. Link to code and data.

5.1. **Time integration with half-explicit Euler.** With the half-explicit Euler time discretisation, at every time step the linear system

$$(5.3) \qquad \begin{bmatrix} \frac{1}{\tau}M + A & B^T \\ B & 0 \end{bmatrix} \begin{bmatrix} v^{k+1} \\ -p^k \end{bmatrix} = \text{rhs} := \begin{bmatrix} \frac{1}{\tau}M - N(v^k) + f \\ g \end{bmatrix}$$

is solved for the velocity and the pressure approximations. For the approximate solution of the linear systems, we use *MinRes* iterations, which are stopped as soon as the relative residual drops below a given tolerance `tol`, i.e.,

$$\left\| \begin{bmatrix} \frac{1}{\tau}M + A & B^T \\ B & 0 \end{bmatrix} \begin{bmatrix} v^{k+1} \\ -p^k \end{bmatrix} - \text{rhs} \right\|_{(M^{-1}, M_{\mathcal{Q}}^{-1})} \leq \text{tol} \, \|\text{rhs}\|_{(M^{-1}, M_{\mathcal{Q}}^{-1})},$$

where we use the norm induced by the inverses of the mass matrix M and of the mass matrix of the pressure approximation space $M_{\mathcal{Q}}$.

It can be seen from the error plots in Figure 5.4, for exact solves, that the half-explicit Euler converges linearly in the velocity and the pressure approximation, which is in line with the theory; see [HLR89, Tab. 2.3] and [BH93]. For inexact solves, however, for smaller time-steps, the pressure approximation diverges linearly. As we have shown in [AH15] this is an inherent instability of the index-2 formulation.

The reported residuals and errors are defined as follows:

$$e^v_{\tau;\text{tol}} = \text{trp}(\|v_{\tau;\text{tol}} - v_{\text{ref}}\|_M)$$

where the subscript `ref` denotes the reference solution, the subscripts $\tau;\text{tol}$ denote the approximation that is computed on the grid of size τ with the linear equations solved with tolerance `tol`, and where $\text{trp}(s)$ denotes the approximation to the integral $\int_0^{t_N} s(t) \, \mathrm{d}t$ by means of the piecewise trapezoidal rule with step size τ. Analogously, we define the error in the pressure approximation

$$e^p_{\tau;\text{tol}} = \text{trp}(\|p_{\tau;\text{tol}} - p_{\text{ref}}\|_{M_P})$$

and the integrated residuals in the momentum equation $r^M_{\tau;\text{tol}}$ as the integral of the function

$$t^k \mapsto \|(\tfrac{1}{\tau}M + A)v^{k+1}_{\tau;\text{tol}} - \tfrac{1}{\tau}Mv^k_{\tau;\text{tol}} - B^T p^k_{\tau;\text{tol}} + N(v^k_{\tau;\text{tol}}) - f^k\|_{M^{-1}}$$

and in the continuum equation as

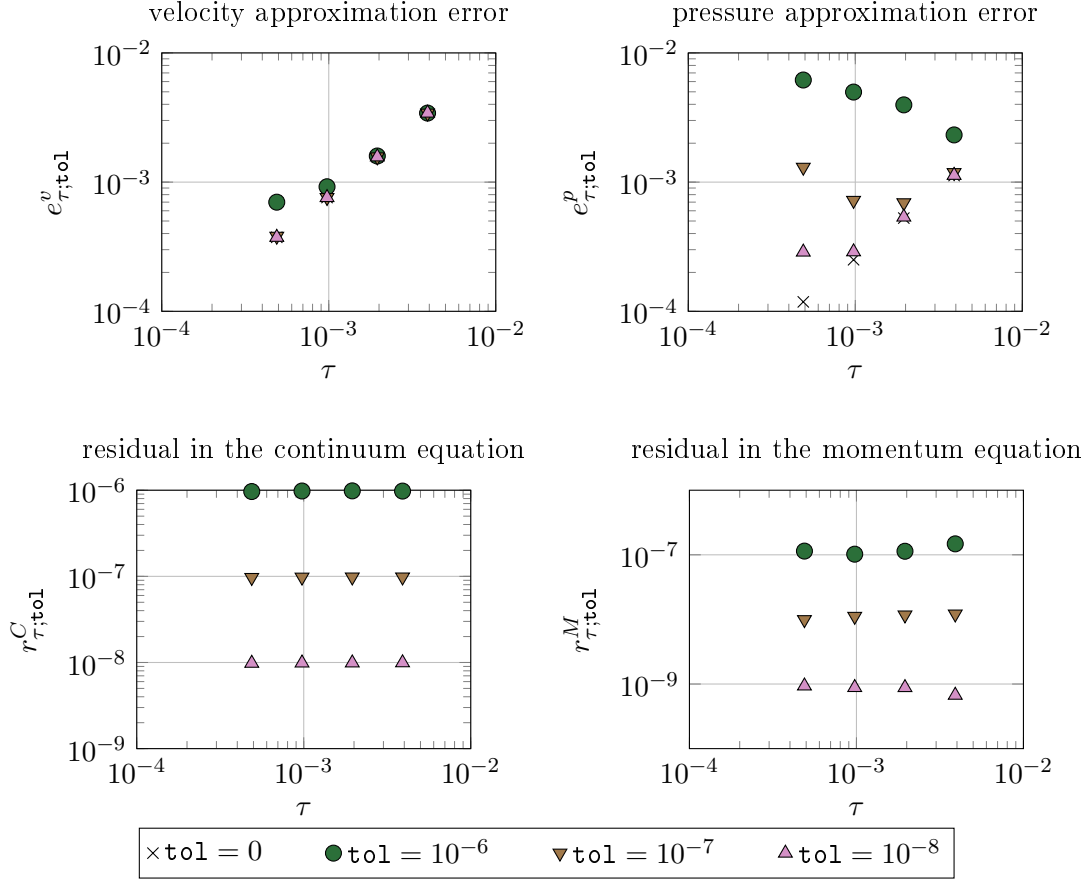$$r^C_{\tau;\text{tol}} := \text{trp}(\|Bv_{\tau;\text{tol}} - g\|_{M_{\mathcal{Q}}^{-1}}).$$

FIGURE 5.4. Error in the velocity and the pressure approximation provided by the half-explicit Euler algorithm for iterative solves with varying tolerances. The crosses are the errors obtained with direct solves.

5.2. **Time integration with SIMPLE.** For this particular time discretisation, we have to solve three linear systems, namely

$$(5.4a) \qquad (\tfrac{1}{\tau}M + A)\tilde{v}^{k+1} = \text{rhs}_1 := \tfrac{1}{\tau}Mv^k + B^Tp^k + f^k,$$

$$(5.4b) \qquad B(\tfrac{1}{\tau}M + A)^{-1}B^Tp_\Delta^{k+1} = \text{rhs}_2 := -B\tilde{v}^{k+1} + g^{k+1},$$

and

$$(5.4c) \qquad (\tfrac{1}{\tau}M + A)v_\Delta^{k+1} = \text{rhs}_3 := B^Tp_\Delta^{k+1}$$

to compute the updates as $v^{k+1} = \tilde{v}^{k+1} + v_\Delta^{k+1}$ and $p^{k+1} = p^k + p_\Delta^{k+1}$. For the approximate solution, we use CG iterations until the relative residuals, measured in the M^{-1} norm (or M_Q^{-1} for (5.4b)), drop below a given tolerance $\texttt{tol}$.

As one can see from the error plots in Figure 5.5, the inexact solves affect the approximation only for a rough tolerance $\texttt{tol} = 10^{-4}$, which is only one order of magnitude smaller than the actual approximation error. Thus the breakdown in the convergence observed in Figure 5.5 is probably due to the accumulation of errors that every single step scheme suffers from. For smaller tolerances, the approximations almost achieve the accuracy of the direct solves. Interestingly, the continuity equation (3.1b), which is only an implicit

velocity approximation error

pressure approximation error

residual in the continuum equation

residual in the momentum equation

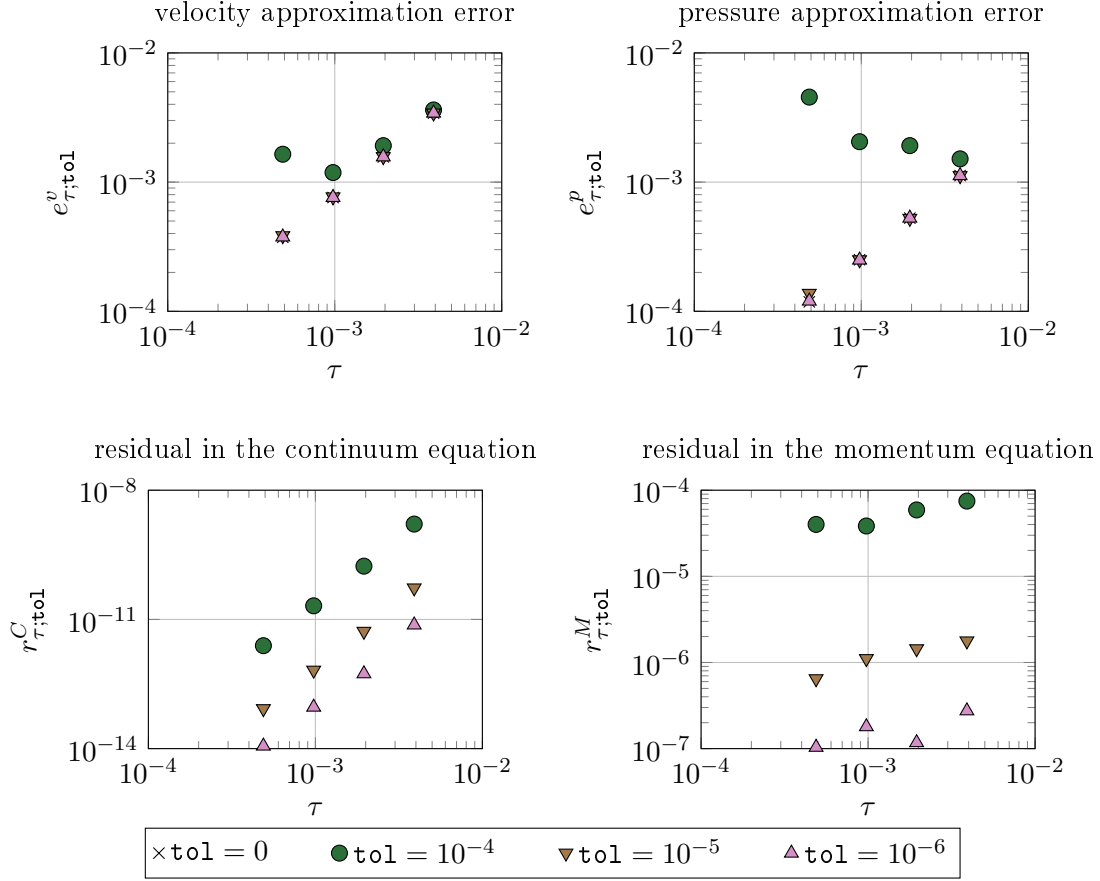| $\times$ tol $= 0$ | $\bullet$ tol $= 10^{-4}$ | $\blacktriangledown$ tol $= 10^{-5}$ | $\blacktriangle$ tol $= 10^{-6}$ |

FIGURE 5.5. Error in the velocity and the pressure approximation provided by the SIMPLE algorithm for exact solves and iterative solves with varying tolerances tol.

part of the SIMPLE scheme, is fulfilled at much better accuracy than the choices of the tolerances suggest; see Figure 5.5.

6. Conclusion

In this paper, we have discussed the incompressible NSE from a DAE point of view, in which the incompressibility is interpreted as an algebraic constraint. Thus, a spatial discretisation leads to a DAE. If the time is discretized as well, a ΔAE is obtained – a sequence of equations, that define the numerical approximations.

We have discussed suitable approaches to well-posed semi-discrete approximations and investigated the ΔAE stemming from different time-discretisations in terms of the Kronecker index. By means of the Kronecker index it became possible to qualitatively compare the time discretized schemes with established numerical schemes that may not have time continuous counterpart. It turned out that commonly and successfully used time integration schemes like the SIMPLE algorithm define a ΔAE of index 1, whereas a time-discretisation of the semi-discrete NSE by an half-explicit Euler scheme leads to a ΔAE of index 2. If, however, an index reduction on the semi-discrete level was applied first, then the time marching schemes led to a ΔAE of lower index.

The advantage of the discrete index-1 formulations is that they avoid implicit derivations that amplify computational errors. This mechanism will likely lead to larger errors, in particular if the equations are solved with limited accuracy. We have illustrated the origin of this behavior in a small analytical example and verified it in a numerical simulation. The code of the numerical test case, an implementation of the 2D cylinder wake, is provided to serve as a benchmark for future developments of time integration schemes for DAEs of Navier-Stokes type.

References

[AH15]   R. Altmann and J. Heiland. Finite element decomposition and minimal extension for flow equations. *ESAIM: M2AN*, 49(5):1489 – 1509, September 2015.

[AH18]   R. Altmann and J. Heiland. Regularization and Rothe discretization of semi-explicit operator DAEs. *Int. J. Numer. Anal. Model.*, 15(3):452–477, 2018.

[ASW93]   M. Arnold, K. Strehmel, and R. Weiner. Half-explicit Runge-Kutta methods for semi-explicit differential-algebraic equations of index 1. *Numerische Mathematik*, 64(1):409–431, Dec 1993.

[BBH17]   M. Behr, P. Benner, and J. Heiland. Example setups of Navier-Stokes equations with control and observation: Spatial discretization and representation via linear-quadratic matrix coefficients. Technical Report arXiv:1707.08711, 2017.

[BF91]   F. Brezzi and M. Fortin. *Mixed and Hybrid Finite Element Methods*. Springer-Verlag, New York, 1991.

[BH93]   V. Brasey and E. Hairer. Half-explicit Runge–Kutta methods for differential-algebraic systems of index 2. *SIAM Journal on Numerical Analysis*, 30(2):538–552, 1993.

[BH15a]   P. Benner and J. Heiland. LQG-Balanced Truncation low-order controller for stabilization of laminar flows. In R. King, editor, *Active Flow and Combustion Control 2014*, pages 365–379. Springer, Berlin, 2015.

[BH15b]   Peter Benner and Jan Heiland. Time-dependent Dirichlet conditions in finite element discretizations. *ScienceOpen Research*, 2015.

[BM14]   M Braack and P. B. Mucha. Directional do-nothing condition for the Navier-Stokes equations. *J. Comput. Math.*, 32(5):507–521, 2014.

[Cam87]   S. Campbell. A general form for solvable linear time varying singular systems of differential equations. *SIAM J. Math. Anal.*, 18(4):1101–1115, 1987.

[Cho68]   A. J. Chorin. Numerical solution of the Navier-Stokes equations. *Math. Comput.*, 22:745–762, 1968.

[Cia78]   P. G. Ciarlet. *The Finite Element Method for Elliptic Problems*. North-Holland, Amsterdam, 1978.

[CM93]   A. J. Chorin and J. E. Marsden. *A Mathematical Introduction to Fluid Mechanics*. Springer-Verlag, New York, third edition, 1993.

[Dai89]   L. Dai. *Singular Control Systems*. Number 118 in Lecture Notes in Control and Information Sciences. Springer, Berlin, 1989.

[EM13]   E. Emmrich and V. Mehrmann. Operator differential-algebraic equations arising in fluid dynamics. *Comp. Methods Appl. Math.*, 13(4):443–470, 2013.

[ESW05]   H. C. Elman, D. J. Silvester, and A. J. Wathen. *Finite Elements and Fast Iterative Solvers: With Applications in Incompressible Fluid Dynamics*. Oxford University Press, Oxford, UK, 2005.

[FK64]   H. Fujita and T. Kato. On the Navier-Stokes initial value problem. i. *Arch. Ration. Mech. Anal.*, 16:269–315, 1964.

[FKP16]   E. Feireisl, T. G. Karper, and M. Pokorný. *Mathematical Theory of Compressible Viscous Fluids. Analysis and Numerics*. Birkhäuser/Springer, Basel, 2016.

[FP02]   J. H. Ferziger and M. Perić. *Computational Methods for Fluid Dynamics*. Berlin: Springer, 3rd edition, 2002.

[Gau17]   A. Gaul. Krypy – a Python toolbox of iterative solvers for linear systems, commit: 36e40e1d. https://github.com/andrenarchy/krypy, 2017.

[GDN97]   M. Griebel, T. Dornseifer, and T. Neunhoeffer. *Numerical Simulation in Fluid Dynamics. A Practical Introduction*. SIAM, Philadelphia, PA, 1997.

[Glo03]   R. Glowinski. Finite element methods for incompressible viscous flow. In *Numerical Methods for Fluids (Part 3)*, volume 9 of *Handbook of Numerical Analysis*, pages 3 – 1176. Elsevier, 2003.

[GR86]   V. Girault and P.-A. Raviart. *Finite Element Methods for Navier–Stokes Equations. Theory and Algorithms*. Springer, Berlin, Germany, 1986.

[Gre90]   P. M. Gresho. On the theory of semi-implicit projection methods for viscous incompressible flow and its implementation via a finite element method that also introduces a nearly consistent mass matrix. I: Theory. *Int. J. Numer. Methods Fluids*, 11(5):587–620, 1990.

[GS00]   P. M. Gresho and R. L. Sani. *Incompressible Flow and the Finite Element Method. Vol. 2: Isothermal Laminar Flow*. Wiley, Chichester, UK, 2000.

[Ha15]   P. Ha. *Analysis and numerical solutions of delay differential algebraic equations*. Phd thesis, Technische Universität Berlin, 2015.

[Hei14]   J. Heiland. *Decoupling and optimization of differential-algebraic equations with application in flow control*. PhD thesis, TU Berlin, 2014. `http://opus4.kobv.de/opus4-tuberlin/frontdoor/index/index/docId/5243`.

[Hin00]   M. Hinze. Optimal and instantaneous control of the instationary Navier-Stokes equations. Habilitationsschrift, Institut für Mathematik, Technische Universität Berlin, 2000.

[HLR89]   E. Hairer, C. Lubich, and M. Roche. *The numerical solution of differential-algebraic systems by Runge-Kutta methods*. Springer-Verlag, Berlin, 1989.

[HLW06]   E. Hairer, C. Lubich, and G. Wanner. *Geometric Numerical Integration. Structure-Preserving Algorithms for Ordinary Differential Equations*. Springer Series in Compu- tational Mathematics. Springer, Berlin, 2 edition, 2006.

[HR90]   J. G. Heywood and R. Rannacher. Finite-element approximation of the nonstationary Navier–Stokes problem. IV: Error analysis for second-order time discretization. *SIAM J. Numer. Anal.*, 27(2):353–384, 1990.

[HV95]   J. C. Heinrich and C. A. Vionnet. The penalty method for the Navier-Stokes equations. *Arch. Comput. Method E.*, 2:51–65, 1995.

[HV16]   X. He and C. Vuik. Comparison of some preconditioners for the incompressible Navier-Stokes equations. *Numerical Mathematics: Theory, Methods and Applications*, 9(2):239–261, 2016.

[HW96]   E. Hairer and G. Wanner. *Solving Ordinary Differential Equations II: Stiff and Differential-Algebraic Problems*. Springer-Verlag, Berlin, second edition, 1996.

[KBN05]   G. Karniadakis, A. Beskok, and A. Narayan. *Microflows and Nanoflows. Fundamentals and Simulation*. Springer, New York, NY, 2005.

[KM94]   P. Kunkel and V. Mehrmann. Canonical forms for linear differential-algebraic equations with variable coefficients. *J. Comput. Appl. Math.*, 56(3):225–251, 1994.

[KM01]   P. Kunkel and V. Mehrmann. Analysis of over- and underdetermined nonlinear differential-algebraic systems with application to nonlinear control problems. *Math. Control Signals Syst.*, 14(3):233–256, 2001.

[KM04]   P. Kunkel and V. Mehrmann. Index reduction for differential-algebraic equations by minimal extension. *Z. Angew. Math. Mech.*, 84(9):579–597, 2004.

[KM06]   P. Kunkel and V. Mehrmann. *Differential-Algebraic Equations. Analysis and Numerical Solution*. European Mathematical Society Publishing House, Zürich, Switzerland, 2006.

[Lad69]   O. A. Ladyzhenskaya. *The Mathematical Theory of Viscous Incompressible Flow*. Gordon and Breach Science Publishers, New York, NY, 1969.

[Lay08]   W. Layton. *Introduction to the Numerical Analysis of Incompressible Viscous Flows*. SIAM, Philadelphia, PA, 2008.

[Ler33]   J. Leray. étude de diverses équations intégrales non linéaires et de quelques problèmes que pose l'hydrodynamique. *J. Math. Pures Appl.*, 12:1–82, 1933.

[LeV92]   R. J. LeVeque. *Numerical Methods for Conservation Laws*. Basel: Birkhäuser, 2nd ed. edition, 1992.

[LL87]   L. D. Landau and E. M. Lifshits. *Fluid Mechanics*, volume 6 of *Course of Theoretical Physics*. 2 edition, 1987. Transl. from the Russian by J. B. Sykes and W. H. Reid.

[LORW12]   A. Logg, K. B. Ølgaard, M. E. Rognes, and G. N. Wells. FFC: the FEniCS form compiler. In *Automated Solution of Differential Equations by the Finite Element Method*, pages 227–238. Springer, Berlin, Germany, 2012.

[MS93]   S. E. Mattsson and G. Söderlind. Index reduction in differential-algebraic equations using dummy derivatives. *SIAM J. Sci. Comput.*, 14(3):677–692, 1993.

[NAM$^+$03] B. R. Noack, K. Afanasiev, M. Morzyński, G. Tadmor, and F. Thiele. A hierarchy of low-dimensional models for the transient and post-transient cylinder wake. *J. Fluid Mech.*, 497:335–363, 2003.

[NR15] P. A. Nguyen and J.-P. Raymond. Boundary stabilization of the Navier–Stokes equations in the case of mixed boundary conditions. *SIAM J. Cont. Optim.*, 53(5):3006–3039, 2015.

[Pir89] O. Pironneau. *Finite element methods for fluids.* John Wiley & Sons, Ltd., Chichester; Masson, Paris, 1989. Translated from the French.

[Ray10] J.-P. Raymond. Feedback stabilization of a fluid-structure model. *SIAM J. Cont. Optim.*, 48(8):5398–5443, 2010.

[Rei06] T. Reis. *Systems Theoretic Aspects of PDAEs and Applications to Electrical Circuits.* Shaker, Aachen, Germany, 2006.

[Rey03] O. Reynolds. *Papers on Mechanical und Physical Subjects. Volume III. The Sub-mechanics of the Universe.* Cambridge University Press, 1903.

[Rou05] T. Roubíček. *Nonlinear Partial Differential Equations with Applications.* Birkhäuser, Basel, Switzerland, 2005.

[Sai09] L. Saint-Raymond. *Hydrodynamic limits of the Boltzmann equation.* Berlin: Springer, 2009.

[She95] J. Shen. On error estimates of the penalty method for unsteady Navier-Stokes equations. *SIAM J. Numer. Anal.*, 32(2):386–403, 1995.

[Ste06] A. Steinbrecher. *Numerical Solution of Quasi-Linear Differential-Algebraic Equations and Industrial Simulation of Multibody Systems.* PhD thesis, Technische Universität Berlin, 2006.

[Tar06] L. Tartar. *An Introduction to Navier–Stokes Equation and Oceanography.* Springer, New York, NY, 2006.

[Tem77] R. Temam. *Navier–Stokes Equations. Theory and Numerical Analysis.* North-Holland, Amsterdam, Netherlands, 1977.

[TH73] 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.

[Tur99] S. Turek. *Efficient Solvers for Incompressible Flow Problems. An Algorithmic and Computational Approach.* Springer, Berlin, Germany, 1999.

[Wei96] J. Weickert. Navier-Stokes equations as a differential-algebraic system. Preprint SFB393/96-08, Technische Universität Chemnitz-Zwickau, 1996.

[Wei97] J. Weickert. *Applications of the theory of differential-algebraic equations to partial differential equations of fluid dynamics.* PhD thesis, Fakultät für Mathematik, Technische Universität Chemnitz, 1997.

[Wil96] C. H. K. Williamson. Vortex dynamics in the cylinder wake. *Annu. Rev. Fluid Mech.*, 28(1):477–539, 1996.

[Zei90] E. Zeidler. *Nonlinear functional analysis and its applications. II/A: Linear monotone operators.* Springer, Berlin, Germany, 1990.

Appendix A. Strangeness index of equation (3.1)

We analyze in detail the strangeness index of the DAE (3.1); see [KM06, Def. 4.4] for the precise definition. Note that we do not ask for any assumptions on the nonlinearity K and that we allow the matrix B to be rank-deficient. This then also implies that the differentiation index of (3.1) equals 2 if it is well-defined, i.e., if B is of full rank.

A.1. **Linear case.** Considering any linearisation of the Navier-Stokes equations, i.e., $K(u) = Ku$ in (3.1), we deal with the matrix pair

$$(E, A) = \left(\begin{bmatrix} M & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} K & B^T \\ B & 0 \end{bmatrix} \right).$$

Following [KM06, Th. 3.11], we can construct a to (E, A) (globally) equivalent pair $(\tilde{E}, \tilde{A})$ of the form

$$\tilde{E} = \left[\begin{array}{cc|c} I_b & & \\ & I_{n-b} & \\ \hline & & 0 \\ & & 0 \end{array}\right], \qquad \tilde{A} = \left[\begin{array}{cc|c} 0 & A_{12} & A_{13} \\ 0 & 0 & A_{23} \\ \hline I_b & & 0 \\ 0 & & 0 \end{array}\right]$$

with $A_{13} \in \mathbb{R}^{b,m}$ being of full rank. Thus, the original system (3.1) is equivalent to a system of the form

$$\dot{x}_1 = A_{12}x_2 + A_{13}x_3 + f_1, \qquad \dot{x}_2 = A_{23}x_3 + f_2, \qquad 0 = x_1 + f_3, \qquad 0 = f_4$$

with dimensions $x_1(t) \in \mathbb{R}^b$, $x_2(t) \in \mathbb{R}^{n-b}$, $x_3(t) \in \mathbb{R}^m$. Since we have a differential and an algebraic equation for x_1 (this causes the 'strangeness'), we use the derivative of $0 = x_1 + f_3$ in order to eliminate $\dot{x}_1$ in the first equation. Hence, we consider the pair $(E_{\mathrm{mod}}, A_{\mathrm{mod}})$ with

$$E_{\mathrm{mod}} = \left[\begin{array}{cc|c} 0 & & \\ & I_{n-b} & \\ \hline & & 0 \\ & & 0 \end{array}\right], \qquad A_{\mathrm{mod}} = \left[\begin{array}{cc|c} 0 & A_{12} & A_{13} \\ 0 & 0 & A_{23} \\ \hline I_b & & 0 \\ 0 & & 0 \end{array}\right].$$

Since A_{13} is of full rank, one can show that system $(E_{\mathrm{mod}}, A_{\mathrm{mod}})$ is strangeness-free, cf. the calculation in [KM06, Th. 3.7]. Since we have obtained a strangeness-free system with only one differentiation, system (3.1) has strangeness index one.

A.2. **Nonlinear case.** The general form of a nonlinear DAE is given by

$$F(t, x, \dot{x}) = 0.$$

In regard of system (3.1) we set $x := [q^T, \ p^T]^T$ and define

$$F(t, x, \dot{x}) := \begin{bmatrix} M\dot{q} - K(q) - B^T p - f \\ -Bq + g \end{bmatrix} = E\dot{x} - A(x) - h$$

with

$$E := \begin{bmatrix} M & 0 \\ 0 & 0 \end{bmatrix}, \qquad A_x := \frac{\partial A(x)}{\partial x} = \begin{bmatrix} K_q & B^T \\ B & 0 \end{bmatrix}.$$

In the sequel we show that (3.1) has strangeness index 1 also in the nonlinear case. For this, we assume that B has full rank such that there are no vanishing equations and the pressure variable is uniquely defined. In the case rank $B = b < m$, we consider the following transformation.

Let $C_0 \in \mathbb{R}^{m,m-b}$ be the matrix of full rank satisfying $B^T C_0 = 0$. Furthermore, $C' \in \mathbb{R}^{m,b}$ defines any matrix such that $C = [C_0 \ C'] \in \mathbb{R}^{m,m}$ is invertible. With this, we obtain the relation

$$B^T C = \begin{bmatrix} B^T C_0 & B^T C' \end{bmatrix} = \begin{bmatrix} 0 & \tilde{B}^T \end{bmatrix}$$

with $\tilde{B} \in \mathbb{R}^{b,n}$ having full rank. With the matrix C in hand, we first introduce the new pressure variable $\tilde{p} := C^{-1}p$. Thus, we consider the the pair $z := [q^T, \tilde{p}^T]^T$. As a second step, we multiply equation (3.1) by the block-diagonal matrix $\mathrm{diag}(I_n, C^T)$ from the left. In total, this yields the equivalent DAE

$$M\dot{q} = K(q) + \begin{bmatrix} 0 & \tilde{B}^T \end{bmatrix}\tilde{p} + f, \qquad \begin{bmatrix} 0 \\ \tilde{B} \end{bmatrix} q = C^T g.$$

Note that the constraint contains $(m - b)$ consistency equations of the form $0 = g_1$. Assuming that system (3.1) is solvable, we suppose that these are in fact vanishing equations. Thus, they have no influence on the index of the system. Furthermore, the first $(m - b)$ components of the transformed pressure $\tilde{p}$ do not influence the system. These components are underdetermined and may be omitted, again without changing the index. Leaving out the underdetermined parts as well as the vanishing equations, we obtain a system of the form (3.1) with a full rank matrix B.

In the sequel, we assume that $\operatorname{rank} B = m$ and show that [KM06, Hyp. 4.2] is satisfied for $\mu = 1$. Note that this hypothesis is not satisfied for $\mu = 0$, i.e., the system is not strangeness-free. We define the matrices

$$M_1 := \begin{bmatrix} E & 0 \\ -A_x & E \end{bmatrix}, \qquad N_1 := \begin{bmatrix} A_x & 0 \\ 0 & 0 \end{bmatrix}.$$

We now pass through the list of points of the hypothesis in [KM06, Hyp. 4.2]:

(1) First, we note that the rank of M_1 equals $2n$ and we set $a := 2(n + m) - 2n = 2m$. Thus, the system contains $2m$ algebraic variables (the pressure and the part of q, which is not divergence-free). Furthermore, we define $Z_2 \in \mathbb{R}^{2(n+m),2m}$ by $Z_2^T M_1 = 0$, i.e.,

$$Z_2 = \left[\begin{array}{c|c} 0 & M^{-1}B^T \\ I_m & 0 \\ \hline 0 & 0 \\ 0 & I_m \end{array}\right].$$

(2) As a second step we define $\hat{A}_2 := Z_2^T N_1 [I_{n+m},\ 0]^T$, which yields

$$\hat{A}_2 = \left[\begin{array}{cc|cc} 0 & I_m & 0 & 0 \\ BM^{-1} & 0 & I_m & 0 \end{array}\right] \begin{bmatrix} K_q & B^T \\ B & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} = \left[\begin{array}{c|c} B & 0 \\ BM^{-1}K_u & BM^{-1}B^T \end{array}\right].$$

This matrix has rank $2m$, since the full-rank property of B implies that $BM^{-1}B^T$ is invertible. We define $d := n - m$ as the number of differential variables and $T_2 \in \mathbb{R}^{n+m,n-m}$ by $\hat{A}_2 T_2 = 0$. Let $C \in \mathbb{R}^{n,n-m}$ be a matrix of full rank with $BC = 0$ and $C_2 := -(BM^{-1}B^T)^{-1}BM^{-1}K_u C \in \mathbb{R}^{m,n-m}$. Then, we set

$$T_2 := \begin{bmatrix} C \\ C_2 \end{bmatrix}.$$

(3) Finally, we compute the rank of ET_2. Since C has full rank, this equals $\operatorname{rank} MC = n - m = d$. The matrix $Z_1^T := [C^T\ 0] \in \mathbb{R}^{n-m,n+m}$ satisfies

$$\operatorname{rank} Z_1^T E T_2 = \operatorname{rank} C^T M C = n - m = d.$$

Thus, the hypothesis in [KM06, Hyp. 4.2] is satisfied for $\mu = 1$, which implies that the nonlinear DAE (3.1) has strangeness index one.

Appendix B. Difference-algebraic equation index of the considered systems

In this appendix, we derive the Kronecker index for the discrete schemes considered in Section 4.3.

B.1. **Half-explicit Euler.** We start with the half-explicit Euler discretisation, that gives a scheme $\mathcal{E}x^{k+1} = \mathcal{A}^k x^k + h^k$ with the matrix pair

$$(\mathcal{E},\mathcal{A}) = (\begin{bmatrix} \frac{1}{\tau}M & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} \frac{1}{\tau}M + A & B^T \\ B & 0 \end{bmatrix})$$

as in (4.4). For sufficiently small τ, due to the definiteness of M and the full-rank property of B, the matrix $\mathcal{A}$ is invertible and thus, the pair $(\mathcal{E},\mathcal{A})$ is regular. Let S denote the matrix $BM^{-1}B^T$. If one applies

$$\begin{bmatrix} M^{-\frac{1}{2}} & B^T S \\ 0 & I \end{bmatrix} \begin{bmatrix} I & 0 \\ BM^{-1} & I \end{bmatrix} \to (\mathcal{E},\mathcal{A}) \leftarrow \begin{bmatrix} M^{-\frac{1}{2}} & 0 \\ -S^{-1}BM^{-1}(\frac{1}{\tau}M + A) & I \end{bmatrix}$$

from the left and the right, one finds that $(\mathcal{E},\mathcal{A})$ is similar to

$$(\begin{bmatrix} \frac{1}{\tau}(I - M^{-\frac{1}{2}}B^T S B M^{-\frac{1}{2}}) & 0 \\ \frac{1}{\tau}B & 0 \end{bmatrix},$$

$$\begin{bmatrix} (I - M^{-\frac{1}{2}}B^T S B M^{-\frac{1}{2}})(\frac{1}{\tau}I + M^{-\frac{1}{2}}AM^{-\frac{1}{2}}) - M^{-\frac{1}{2}}B^T S B M^{-\frac{1}{2}} & 0 \\ 0 & S \end{bmatrix}).$$

Since B is of full rank, there exists an orthogonal matrix Q and an invertible matrix R such that $BM^{-\frac{1}{2}}Q = \begin{bmatrix} 0 & R \end{bmatrix}$ and, in particular,

$$Q^T(I - M^{-\frac{1}{2}}B^T S B M^{-\frac{1}{2}})Q = \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix}.$$

Thus, the corresponding similarity transformation transforms $(\mathcal{E},\mathcal{A})$ into

$$(\begin{bmatrix} \frac{1}{\tau}I & 0 & 0 \\ 0 & 0 & 0 \\ 0 & \frac{1}{\tau}R & 0 \end{bmatrix}, \begin{bmatrix} \tilde{a}_{11} & 0 & 0 \\ \tilde{a}_{21} & I & 0 \\ 0 & 0 & S \end{bmatrix}),$$

where $\tilde{a}_{11}$ and $\tilde{a}_{21}$ stand for unspecified but possibly nonzero block matrix entries. With another few regular row and column transformations, one can eliminate the entry $\tilde{a}_{21}$ and read off the Kronecker index of $(\mathcal{E},\mathcal{A})$ as the index of nilpotency of $\begin{bmatrix} 0 & 0 \\ \frac{1}{\tau}R & 0 \end{bmatrix}$ which is 2.

B.2. **Projection scheme.** The matrix coefficient pair of the Projection scheme (4.9) reads

(B.1) $$(\mathcal{E},\mathcal{A}) = (\begin{bmatrix} \frac{1}{\tau}M & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ -M & -\frac{\tau}{2}B^T & M & 0 \\ 0 & -I & 0 & I \end{bmatrix}, \begin{bmatrix} 0 & \frac{1}{\tau}M + A & 0 & B^T \\ -\frac{2}{\tau}B & -BM^{-1}B^T & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & I \end{bmatrix}).$$

If we define $S := BM^{-1}B^T$, if we move the second row and column to the left and bottom, respectively, and if we rescale certain rows and columns, we find that the pair is equivalent to

$$(\mathcal{E},\mathcal{A}) \sim (\begin{bmatrix} I & 0 & 0 & 0 \\ \tilde{e}_{21} & I & 0 & \tilde{e}_{24} \\ 0 & 0 & I & \tilde{e}_{34} \\ 0 & 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & \tilde{a}_{13} & \tilde{a}_{14} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & \tilde{a}_{33} & 0 \\ \tilde{a}_{41} & 0 & 0 & -S \end{bmatrix}),$$

where the $\tilde{e}$'s and $\tilde{a}$'s stand for unspecified but possibly nonzero entries. Since, in particular, S is invertible, one can eliminate the entries $\tilde{e}_{24}$, $\tilde{e}_{34}$, $\tilde{a}_{41}$, and $\tilde{a}_{14}$ by regular row and column manipulations without affecting the invertibility of the left upper 3×3 block in the transformed $\mathcal{E}$ and read off the Kronecker index of (4.9) as the index of nilpotency of 0 which is 1.

B.3. **SIMPLE.** The matrix coefficient pair of the SIMPLE scheme (4.13) reads

$$\left(\mathcal{E}, \mathcal{A}\right) = \left(\begin{bmatrix} \frac{1}{\tau}M & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ I & (\frac{1}{\tau}M + A)^{-1}B^T & -I & 0 \\ 0 & -I & 0 & I \end{bmatrix}, \begin{bmatrix} 0 & \frac{1}{\tau}M + A & 0 & B^T \\ -B & -B(\frac{1}{\tau}M + A)^{-1}B^T & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & I \end{bmatrix} \right).$$

If we define $S_A := B(\frac{1}{\tau}M^{-1} + A)^{-1}B^T$, move the second row and column to the left and bottom, respectively, and rescale certain rows and columns, then we find that the pair is equivalent to

$$\left(\mathcal{E}, \mathcal{A}\right) \sim \left(\begin{bmatrix} I & 0 & 0 & 0 \\ \tilde{e}_{21} & I & 0 & \tilde{e}_{24} \\ 0 & 0 & I & \tilde{e}_{34} \\ 0 & 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & \tilde{a}_{13} & \tilde{a}_{14} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & \tilde{a}_{33} & 0 \\ \tilde{a}_{41} & 0 & 0 & -S_A \end{bmatrix} \right),$$

where, again, the $\tilde{e}$'s and $\tilde{a}$'s stand for unspecified but possibly nonzero entries. Since S_A is invertible for sufficiently small τ, we find that this matrix pair has the very same structure as the one of the projection scheme (see Appendix B.2) and, thus, is of index 1.

* Department of Mathematics, University of Augsburg, Universitätsstr. 14, 86159 Augsburg, Germany

† Max Planck Institute for Dynamics of Complex Technical Systems, Sandtorstrasse 1, 39106 Magdeburg, Germany

E-mail address: robert.altmann@math.uni-augsburg.de, heiland@mpi-magdeburg.mpg.de