INTERNATIONAL JOURNAL OF
NUMERICAL ANALYSIS AND MODELING
Volume 15, Number 3, Pages 452–478

REGULARIZATION AND ROTHE DISCRETIZATION OF SEMI-EXPLICIT OPERATOR DAES

ROBERT ALTMANN AND JAN HEILAND

Abstract. A general framework for the regularization of constrained PDEs, also called operator differential-algebraic equations (operator DAEs), is presented. The given procedure works for semi-explicit and semi-linear operator DAEs of first order including the Navier-Stokes and other flow equations. The proposed reformulation is consistent, i.e., the solution of the PDE remains untouched. Its main advantage is that it regularizes the operator DAE in the sense that a semi-discretization in space leads to a DAE of lower index. Furthermore, a stability analysis is presented for the linear case, which shows that the regularization provides benefits also for the application of the Rothe method. For this, the influence of perturbations is analyzed for the different formulations. The results are verified by means of a numerical example with an adaptive space discretization.

Key words. PDAE, operator DAE, regularization, index reduction, Rothe method, method of lines, perturbation analysis

1. Introduction

Constrained PDEs arise naturally in the modelling of physical, chemical, and many other real-world phenomena. They occur whenever different PDE models are coupled, e.g., via mutual variables at the interfaces, since the coupling is typically modelled via algebraic constraints. Such models are widely used in flexible multibody dynamics, e.g., the pantograph and catenary benchmark problem [6] or the flexible slider crank mechanism [30, 31]. Also flow equations such as the Navier-Stokes equations [33, 35] can be seen as constrained PDEs due to the coupling of momentum equation to the divergence-free constraint. Further applications can be found in circuit simulation [34], electromagnetics, and chemical engineering [9].

We consider these equation systems of ordinary or partial differential equations (ODEs, PDEs) and algebraic equations in line with other constrained PDEs – often referred to as PDAEs – as differential-algebraic equations (DAEs) in function spaces, so-called *abstract* or *operator DAEs*.

Despite the large range of applications and the advantages from the modeling perspective, the mathematical analysis of operator DAEs is full of open research questions. There is still no common classification like the index concepts for DAEs [21, Ch. 12]. The generalization of the *tractability index* as proposed, e.g., in [34] does not apply for the commonly used formulation by means of Gelfand triples. The very general concept of the *perturbation index*, as it was defined in [25] for linear PDAEs, applies under strong regularity conditions but is still ambiguous in the choice of the norm in which one measures the perturbation and their derivatives. Also the *differentiation index* was generalized to PDAEs [22] but has difficulties with the agreement of the PDAE index with the index of the semi-discretized DAE. Yet another idea is to classify the index of a PDAE directly by the index that may

Received by the editors April 30, 2016.
2000 *Mathematics Subject Classification.* 65J08, 65M12, 65L80.

be determined after a spatial discretization. This, however, leads to the similar unclear problem, what a good discretization of a PDAE is.

Within this paper, we analyse constrained systems of first order and semi-explicit structure. Particularly, we consider systems of the form

$$\dot{u}(t) + \mathcal{K}u(t) + \mathcal{B}^*\lambda(t) = \mathcal{F}(t), \tag{1a}$$

$$\mathcal{B}u(t) \qquad\qquad = \mathcal{G}(t). \tag{1b}$$

Therein, λ denotes the Lagrange multiplier, which enforces the linear constraint $\mathcal{B}u = \mathcal{G}$. In view of numerical simulations, the incorporation of the constraints via a Lagrange multiplier and a suitable reformulation seem promising and follow the paradigm in the treatment of DAEs, that it is preferable to collect all available information in the form of constraints instead of eliminating them. For the Navier-Stokes equations this means to maintain the pressure as part of the system.

In this paper, we introduce a regularization or *index reduction method* for the PDAE (1) without introducing an index as such. We rather refer to the well-defined index of the semi-discrete system after a spatial discretization by mixed finite elements. In other words, we propose a reformulation of the given PDAE system such that a semi-discretization leads to a DAE of lower index.

A transformation on operator level can be the base for numerically advantageous discretization schemes. The commonly taken approach of first discretizing and then transforming the equations comes with the latent risk that the algebraic manipulation are not valid in infinite dimensions [17]. This may cause instabilities or inconsistencies as the discretization becomes more accurate. The taken approach has been introduced for second-order systems appearing in elastodynamics [2] and for flow equations [4] before. Here, we consider the more general case with time-dependent constraints and provide the functional analytical framework.

The main contribution of this paper is then the analysis of the Rothe discretization through the application of the implicit Euler scheme to the operator DAE (1). As for the finite-dimensional case, we expect a different behavior of the variables u and λ. It will turn out that we need stronger regularity assumptions to prove the convergence of the Lagrange multiplier. Among others, we consider the influence of perturbations and quantify them in a general convergence result. We show that the proposed reformulation improves the robustness against such perturbations, as we confirm numerically for a simulation setup with adaptive, and thus changing, meshes.

The paper is organized as follows. In Section 2 we provide the theoretical framework for the formulation of operator differential equations. These tools are then used for the formulation and regularization of the operator DAEs in Section 3 in which we also analyse the influence of perturbations. The advantages of the obtained formulation is topic of Section 4. We consider the discretization in time, which corresponds to the Rothe method for time-dependent PDEs in Section 5. Further, we prove the convergence of the implicit Euler scheme and discuss the resulting advantages in terms of perturbations. Finally, we illustrate the obtained theoretical results in a numerical simulation of the Navier-Stokes equations in Section 6 and conclude the paper in Section 7.

2. Preliminaries

This section is devoted to the introduction of the functional analytical backgroud as well as some basics on DAEs. Both ingredients are necessary to understand the notion of operator DAEs. First, we discuss the spaces and operators, which are needed for the analysis in Sections 3 and 5 below. Throughout this paper, we use the notion of Sobolev spaces as in [1] and Bochner spaces as in [28, Ch. 1.5]. Second, we give the definition of the differentiation index and introduce the idea of *minimal extension*, on which the regularization in Section 3.2 is based.

2.1. Sobolev-Bochner spaces. To keep the setting as general as possible, we consider a real, separable, and reflexive Banach space $\mathcal{V}$ and a real separable Hilbert space $\mathcal{H}$ with inner product $(\cdot,\cdot)$. We assume that the spaces $\mathcal{V}$, $\mathcal{H}$, and $\mathcal{V}^*$ form a *Gelfand triple* (also called evolution triple) [36, Ch. 23.4]. This means that $\mathcal{V}$ is densely, continuously embedded in $\mathcal{H}$, written as $\mathcal{V} \hookrightarrow \mathcal{H}$, and that $\mathcal{H}$ and its dual space $\mathcal{H}^*$ are identified via the *Riesz isomorphism*. Such a triple implies the inclusion $\mathcal{H}^* \hookrightarrow \mathcal{V}^*$ in the sense that for $h \in \mathcal{H} \cong \mathcal{H}^*$ and $v \in \mathcal{V}$ we have

$$\langle h, v\rangle_{\mathcal{V}^*,\mathcal{V}} = (h, v).$$

The space for the Lagrange multiplier is denoted by $\mathcal{Q}$ and is assumed to be a real, separable, and reflexive Banach space. The constraint operator $\mathcal{B}$ then maps from $\mathcal{V}$ to $\mathcal{Q}^*$. Together with its dual operator $\mathcal{B}^*$, we obtain the following diagram:

$$\begin{array}{ccccc} \mathcal{V} & \hookrightarrow & \mathcal{H} = \mathcal{H}^* & \hookrightarrow & \mathcal{V}^* \\ {\scriptstyle\mathcal{B}}\downarrow & & & & \uparrow{\scriptstyle\mathcal{B}^*} \\ \mathcal{Q}^* & & & & \mathcal{Q} \end{array}$$

Example 2.1. A typical example for a Gelfand triple $\mathcal{V} \hookrightarrow \mathcal{H} \hookrightarrow \mathcal{V}^*$ is given by the Sobolev spaces $\mathcal{V} := H_0^1(\Omega)$, $\mathcal{H} := L^2(\Omega)$, and $\mathcal{V}^* = H^{-1}(\Omega)$.

We consider time derivatives in the generalized sense as defined, e.g., in [36, Ch. 23.5]. We require solutions of system (1) to satisfy

$$u \in L^p(0,T;\mathcal{V}) \quad \text{with} \quad \dot{u} \in L^q(0,T;\mathcal{V}^*),$$

where $1 < q \le p < \infty$. If q is the conjugated exponent, i.e., $1/p + 1/q = 1$, then, by the well-known embedding theorems for Gelfand triples [36, Th. 23.23], it holds that such a solution u is continuous as a function $u\colon [0,T] \to \mathcal{H}$, i.e., $u \in C([0,T],\mathcal{H})$. Thus, an initial condition $u(0) = a$ for $a \in \mathcal{H}$ is well-defined.

Remark 2.1. The regularization proposed in Section 3 operates with splittings of the state space $\mathcal{V}$ and is independent of the time regularity of the function u or $\dot{u}$. Thus, we can also consider less regular systems with $\dot{u} \in L^q(0,T;\mathcal{V}^*)$ with $q \le 1 - 1/p$, as they may appear in applications. However, we will have to assume the well-posedness of the initial condition in this case.

2.2. Operator $\mathcal{K}$. Consider a possibly nonlinear operator $\mathcal{K}\colon (0,T) \times \mathcal{V} \to \mathcal{V}^*$ and $1 \le q,\ p < \infty$. The question arises whether this operator induces a (bounded) operator of the form

$$\begin{aligned} \mathcal{K}\colon L^p(0,T;\mathcal{V}) &\to L^q(0,T;\mathcal{V}^*), \\ (\mathcal{K}u)(t) &:= \mathcal{K}(t,u(t)). \end{aligned}$$

If such an operator exists, then we do not distinguish between these two notions. We state a well-known result for *Nemytskij* mappings for the considered setup of abstract functions.

Theorem 2.1 (cf. [28, Thm. 1.43])**.** *If the operator* $\mathcal{K}\colon (0,T)\times\mathcal{V}\to\mathcal{V}^*$ *is such that*

(a) $\mathcal{K}(t,\cdot)\colon \mathcal{V}\to\mathcal{V}^*$ *is continuous for almost all* $t\in(0,T)$,
(b) $\mathcal{K}(\cdot,v)\colon (0,T)\to\mathcal{V}^*$ *is measurable for all* v, *and*
(c) $\|\mathcal{K}(t,v)\|_{\mathcal{V}^*}\le\gamma(t)+c\|v\|_{\mathcal{V}}^{p/q}$ *for some* $\gamma\in L^q(0,T)$,

then the mapping defined via

$$(\mathcal{K}v)(t):=\mathcal{K}(t,v(t)),$$

is continuous as a map $\mathcal{K}\colon L^p(0,T;V)\to L^q(0,T;V^*)$, *where* $1\le p<\infty$ *and* $1\le q\le\infty$.

The case that the exponents $1<p,\,q<\infty$ are conjugated, i.e, $1/p+1/q=1$, is often assumed for the analysis of nonlinear evolution equations with monotonicity arguments [28, Ch. 2 and Ch. 8]. However, for nonlinear operators, even if they are uniformly bounded as a map $\mathcal{V}\to\mathcal{V}^*$, the conjugacy of the time exponents may not hold a priori [12, Ch. 8.2].

Example 2.2 (Navier-Stokes operator)**.** Consider the nonlinear operator, which arises in the weak formulation of the Navier-Stokes equations,

$$\mathcal{K}\colon \mathcal{V}\to\mathcal{V}^*,\quad \langle\mathcal{K}u,w\rangle_{\mathcal{V}^*,\mathcal{V}}:=\int\big(u\cdot\nabla\big)\cdot uw\,\mathrm{d}x.$$

Then, $\mathcal{K}\colon\mathcal{V}\to\mathcal{V}^*$ is bounded independently of t, cf. [33, Lem. II.1.1], but, in the three-dimensional case, it is only bounded as an operator $\mathcal{K}\colon L^2(0,T;\mathcal{V})\cap L^\infty(0,T;\mathcal{H})\to L^{4/3}(0,T;\mathcal{V}^*)$, see e.g. [28, Ch. 8.8.4].

Example 2.3 (p-Laplacian)**.** For the p-Laplacian, i.e.,

$$\big\langle\mathcal{K}u,v\big\rangle_{\mathcal{V}^*,\mathcal{V}}:=\int_\Omega|\nabla u|^{p-2}\nabla u\cdot\nabla v\,\mathrm{d}x,$$

we take the Sobolev space $\mathcal{V}=W_0^{1,p}(\Omega)$. This then induces an operator $\mathcal{K}\colon L^p\big(0,T;\mathcal{V}\big)\to L^{p'}(0,T;\mathcal{V}^*)$ with $1/p+1/p'=1$, see [29, Ch. 3.3.6].

For special operators $\mathcal{K}$, as, e.g., linear operators that are uniformly bounded with respect to time, we state the following result.

Corollary 2.2. *Consider* $1\le p<\infty$ *and an operator* $\mathcal{K}\colon(0,T)\times\mathcal{V}\to\mathcal{V}^*$, *which is measurable for fixed* $v\in\mathcal{V}$ *and uniformly bounded in the sense that there exists a constant* $C_{\mathcal{K}}$ *such that* $\|\mathcal{K}(t)v\|_{\mathcal{V}^*}\le C_{\mathcal{K}}\|v\|_{\mathcal{V}}$ *for all* $v\in\mathcal{V}$ *and almost all* $t\in(0,T)$. *Then,* $(\mathcal{K}v)(t):=\mathcal{K}(t,v(t))$ *defines a continuous operator from* $L^p(0,T;\mathcal{V})$ *to* $L^p(0,T;\mathcal{V}^*)$.

Proof. The application of Theorem 2.1 with $p=q$ and $\gamma=0$ yields the result. □

2.3. Differential-algebraic equations. As mentioned in the introduction, differential-algebraic equations (DAEs) are commonly classified through an *index*, which can also be seen as a measure for the expected difficulty of solving such a system numerically. We emphasize that there exist several different index concepts [24] but confine ourselves to the so-called *differentiation index*. Furthermore, we restrict our considerations to semi-explicit systems of the form

$$\dot{q}(t) = f(t, q(t), \mu(t)), \qquad 0 = g(t, q(t)) \tag{2}$$

with $q(t) \in \mathbb{R}^n$ and $\mu(t) \in \mathbb{R}^m$, $m \leq n$. This means that the constraint for the differential variable q is explicitly given by the function $g\colon [0,T] \times \mathbb{R}^n \to \mathbb{R}^m$.

The differentiation index quantifies the necessary number of differentiation steps in order to obtain an ODE and thus, describes to which degree the solution depends on derivatives of the involved quantities. Note that the dependence on derivatives may lead to instabilities within the numerical simulation.

Definition 2.1 ([8, Def. 2.2.2])**.** The minimal number of times that all or parts of (2) must be differentiated with respect to time t in order to determine $\dot{q}$ and $\dot{\mu}$ as continuous functions of q, μ, and t is called the *differentiation index*.

For semi-explicit systems it is known to be sufficient to consider differentiations of the constraint. This then leads to the following result.

Lemma 2.3 ([16, Ch. VII.1])**.** *The semi-explicit DAE* (2) *has differentiation index 2 if the matrix* $\frac{\partial g}{\partial q}\frac{\partial f}{\partial \mu}$ *is invertible.*

We emphasize that DAEs of index 1 can, in principle, be numerically treated as stiff ODEs [16, Ch. VI.1]. For DAEs of higher index, however, one may observe a reduction of the convergence order or even a loss of convergence [24]. Thus, a direct treatment is not advisable [8, Ch. 5.4]. A better approach is to reformulate the system such that the solution set remains unchanged but the index is reduced. Such methods are called *index reduction*.

One particular index reduction method, which is well suited for semi-explicit systems of the form (2), is called *minimal extension*, cf. [19, 23]. We will adapt this technique for the operator case in Section 3.2. For an introduction of this method we consider a special case of (2), namely

$$M\dot{q} = f(t, q, \mu) = \tilde{f}(t, q) - G^T\mu, \qquad 0 = g(t, q)$$

with a positive definite mass matrix $M \in \mathbb{R}^{n,n}$ and the Jacobian $G := \partial g/\partial q \in \mathbb{R}^{m,n}$, which is assumed to be of full row rank $m \leq n$. Applying M^{-1} from the left, we obtain by Lemma 2.3 that this DAE is of index 2, since $\frac{\partial g}{\partial q}\frac{\partial f}{\partial \mu} = GM^{-1}G^T$ is invertible. Because of the full rank property, there exists an orthogonal matrix $Q \in \mathbb{R}^{n,n}$ such that GQ has the block structure $GQ = [G_1,\ G_2]$ with an invertible matrix $G_2 \in \mathbb{R}^{m,m}$. Accordingly, we transform the variable q into

$$\begin{bmatrix} q_1 \\ q_2 \end{bmatrix} := Q^T q$$

with $q_1 \in \mathbb{R}^{n-m}$ and $q_2 \in \mathbb{R}^m$. With this, we can write the derivative of the constraint as

$$0 = \frac{\mathrm{d}}{\mathrm{d}t} g(q) = G\dot{q} + \partial_t g(q).$$

As the name of the method reveals, we extend the system and introduce a dummy variable. Replacing the derivative of q_2 by $\tilde{q}_2 := \dot{q}_2$, we obtain the system

$$\begin{aligned} MQ \begin{bmatrix} \dot{q}_1 \\ \tilde{q}_2 \end{bmatrix} &= \tilde{f}(t, q_1, q_2) - G^T \mu, \\ 0 &= g(t, q_1, q_2), \\ 0 &= G_1 \dot{q}_1 + G_2 \tilde{q}_2 + \partial_t g(t, q_1, q_2), \end{aligned}$$

which is equivalent to the original system but, as one can show, of index 1.

3. Regularization of operator DAEs

In this section, we consider semi-explicit operator equations in a time interval $(0, T)$. With a given linear constraint incorporated by the *Lagrangian method*, we obtain a system of the form: find $u\colon (0,T) \to \mathcal{V}$ and $\lambda\colon (0,T) \to \mathcal{Q}$ such that

$$\dot{u}(t) + \mathcal{K}(t)u(t) + \mathcal{B}(t)^*\lambda(t) = \mathcal{F}(t) \quad \text{in } \mathcal{V}^*, \tag{3a}$$

$$\mathcal{B}(t)u(t) \qquad\qquad\quad = \mathcal{G}(t) \quad \text{in } \mathcal{Q}^* \tag{3b}$$

for $t \in (0,T)$ a.e. with initial condition

$$u(0) = a \in \mathcal{H}. \tag{3c}$$

Therein, $\mathcal{B}^*(t)$ denotes the dual of the linear constraint operator $\mathcal{B}(t)$. System (3) is a generalization of a semi-explicit DAE since here, $u(t)$ belongs to the infinite-dimensional Banach space $\mathcal{V}$. Because of this, we call system (3) a semi-explicit *operator DAE*.

Suitable function spaces for the solution (u, λ) will be discussed in Theorem 3.4 below. We assume $\mathcal{F} \in L^q(0,T;\mathcal{V}^*)$ and $\mathcal{G} \in L^p(0,T;\mathcal{Q}^*)$. The equalities (3a) and (3b) should be understood pointwise in L^1_{loc} in the corresponding dual product. By the *fundamental theorem of variational calculus* [12, Thm. 8.1.3] and the definition of the weak time derivative, this means that $\dot{v}(t) = \mathcal{F}(t)$ in $\mathcal{V}^*$ if

$$-\int_0^T \langle v(t), w \rangle_{\mathcal{V}^*,\mathcal{V}} \dot{\phi}(t)\,\mathrm{d}t = \int_0^T \langle \mathcal{F}(t), w \rangle_{\mathcal{V}^*,\mathcal{V}} \phi(t)\,\mathrm{d}t$$

for all $w \in \mathcal{V}$ and $\phi \in \mathcal{C}_0^\infty(0,T)$. Furthermore, we assume operators $\mathcal{K}\colon L^p(0,T;\mathcal{V}) \to L^q(0,T;\mathcal{V}^*)$, cf. Section 2.2, and $\mathcal{B}\colon L^p(0,T;\mathcal{V}) \to L^p(0,T;\mathcal{Q}^*)$ with $1 < p \le q < \infty$.

3.1. Assumptions on $\mathcal{B}$. In this subsection, we summarize the properties of the constraint operator $\mathcal{B}$, which we require for a reformulation of the operator DAE (3). Note that we do not need additional assumptions of $\mathcal{K}$ at this point.

Assumption 3.1 (Properties of $\mathcal{B}$). The constraint operator $\mathcal{B}(t)\colon \mathcal{V} \to \mathcal{Q}^*$ satisfies the following conditions:

(a) $\mathcal{B}(t)$ is linear and uniformly bounded, $\mathcal{B}(\cdot)v$ is measurable for all $v \in \mathcal{V}$,
(b) $\mathcal{V}_\mathcal{B} := \ker \mathcal{B}(t)$ is independent of time t,
(c) there exists a uniformly bounded right-inverse of $\mathcal{B}(t)$, i.e., there exists a uniformly bounded operator $\mathcal{B}^-(t)\colon \mathcal{Q}^* \to \mathcal{V}$ such that for all $q \in \mathcal{Q}^*$ it holds that
$$\mathcal{B}(t)\mathcal{B}^-(t)q = q,$$
(d) the range of the right-inverse $\mathcal{V}^c := \operatorname{range} \mathcal{B}^-(t)$ is independent of time t,
(e) there exist continuous time derivatives $\dot{\mathcal{B}}(t)\colon \mathcal{V} \to \mathcal{Q}$ and $\dot{\mathcal{B}}^-(t)\colon \mathcal{Q}^* \to \mathcal{V}$.

Remark 3.1 (Time-independent constraint). If the constraint operator is independent of time, i.e., $\mathcal{B}(t) \equiv \mathcal{B}$, then Assumption 3.1 reduces to the points (a) and (c).

Remark 3.2 (Induced operators). By Corollary 2.2 it follows that $\mathcal{B}(t)$ and $\mathcal{B}^-(t)$ from Assumption 3.1 induce bounded operators of the form

$$\mathcal{B}\colon L^p(0,T;\mathcal{V}) \to L^p(0,T;\mathcal{Q}^*) \quad \text{and} \quad \mathcal{B}^-\colon L^p(0,T;\mathcal{Q}^*) \to L^p(0,T;\mathcal{V}^c).$$

Example 3.1. In the intended application of flow problems, the constraint operator will be the divergence operator which, in the common weak formulation and under standard assumptions on the physical or computational domain, will fulfill Assumption 3.1, cf. Section 6.1 below.

If the domain Ω changes with time, as it happens in models of fluid-structure interaction problems, then the divergence operator will depend on time. In principle, Assumption 3.1 and the following results permit such a time dependency but the validity of the assumption will be problem specific.

An example of a time-dependent constraint operator, that is directly covered by Assumption 3.1, would be a time-dependent linear combination of constraint operators, i.e.,

$$\mathcal{B}(t)u = \alpha_1(t)\mathcal{B}_1 u + \alpha_2(t)\mathcal{B}_2 u$$

with $\alpha_1(t), \alpha_2(t) > 0$ for all t and time-independent linear and bounded operators $\mathcal{B}_1, \mathcal{B}_2$.

Note that the choice of the right-inverse in Assumption 3.1 is not unique. A special case, for which the existence of a right-inverse is guaranteed, is when $\mathcal{B}(t)$ satisfies an inf-sup condition of the form

$$\inf_{q\in\mathcal{Q}} \sup_{v\in\mathcal{V}} \frac{\langle \mathcal{B}(t)v, q\rangle}{\|v\|_{\mathcal{V}}\|q\|_{\mathcal{Q}}} \geq \beta > 0.$$

Nevertheless, this does not imply the time-independence of the range of $\mathcal{B}^-(t)$. In the next lemma, we summarize several properties of the right-inverse $\mathcal{B}^-(t)$ from Assumption 3.1.

Lemma 3.1 (Properties of $\mathcal{B}^-$)**.** *Let $\mathcal{B}$ satisfy Assumption 3.1. Then, the right-inverse $\mathcal{B}^-(t)\colon \mathcal{Q}^* \to \mathcal{V}$ is linear and one-to-one. Furthermore, $\mathcal{V}^c := \operatorname{range}\mathcal{B}^-(t)$ is a closed subspace of $\mathcal{V}$ and the operator $\mathcal{B}^-(t)\mathcal{B}(t)\colon \mathcal{V} \to \mathcal{V}$, restricted to $\mathcal{V}^c$, equals the identity.*

Proof. The linearity of $\mathcal{B}^-(t)$ follows from the linearity of the operator $\mathcal{B}(t)$ [26, Ch. 8.1.2]. For the one-to-one relation, consider $q_1,\ q_2 \in \mathcal{Q}^*$ with $\mathcal{B}^-(t)q_1 = \mathcal{B}^-(t)q_2$. Then, the application of $\mathcal{B}(t)$ yields $q_1 = \mathcal{B}(t)\mathcal{B}^-(t)q_1 = \mathcal{B}(t)\mathcal{B}^-(t)q_2 = q_2$.

The linearity of $\mathcal{B}^-(t)$ and the continuity of $\mathcal{B}^-(t)$ and $\mathcal{B}(t)$ imply that $\mathcal{V}^c$ is a closed subspace of $\mathcal{V}$. Finally, for $v \in \mathcal{V}^c$ and fixed $t \in (0,T)$ there exists $q \in \mathcal{Q}^*$ with $\mathcal{B}^-(t)q = v$. Then, Assumption 3.1 implies

$$v = \mathcal{B}^-(t)q = \mathcal{B}^-(t)\big(\mathcal{B}(t)\mathcal{B}^-(t)q\big) = \mathcal{B}^-(t)\mathcal{B}(t)v. \qquad \square$$

Remark 3.3. Lemma 3.1 implies that $\mathcal{B}^-(t)\mathcal{B}(t)\colon \mathcal{V} \to \mathcal{V}$ is a projection onto $\mathcal{V}^c$. Furthermore, the induced Nemytskij mapping $\mathcal{B}^-$ is a right-inverse of $\mathcal{B}$.

An important implication of Assumption 3.1 and Lemma 3.1 is the decomposition of $L^p(0,T;\mathcal{V})$ as given in the following lemma. This decomposition will be the basis for the index reduction procedure of Section 3.2.

Lemma 3.2 (Decomposition of $L^p(0,T;\mathcal{V})$)**.** *Consider the subspaces $\mathcal{V}_\mathcal{B}$ and $\mathcal{V}^c$ of $\mathcal{V}$ defined in Assumption 3.1. Then, we have the decomposition*

$$L^p(0,T;\mathcal{V}) = L^p(0,T;\mathcal{V}_\mathcal{B}) \oplus L^p(0,T;\mathcal{V}^c).$$

Proof. For given $v \in L^p(0,T;\mathcal{V})$, we define $r := \mathcal{B}v \in L^p(0,T;\mathcal{Q}^*)$, cf. Remark 3.2. A decomposition of $v \in L^p(0,T;\mathcal{V})$ is then given by

$$v = v_0 + v^c := \big(v - \mathcal{B}^- r\big) + \mathcal{B}^- r. \tag{4}$$

Obviously, $v^c = \mathcal{B}^- r \in L^p(0,T;\mathcal{V}^c)$ and $v_0 \in L^p(0,T;\mathcal{V}_\mathcal{B})$ follows from Assumption 3.1 by $\mathcal{B}v_0 = \mathcal{B}v - \mathcal{B}\mathcal{B}^-\mathcal{B}v = 0$. We show that the decomposition in (4) is unique. For this, consider v_0, $w_0 \in L^p(0,T;\mathcal{V}_\mathcal{B})$ and v^c, $w^c \in L^p(0,T;\mathcal{V}^c)$ with $v = v_0 + v^c = w_0 + w^c$. The application of $\mathcal{B}$ yields $\mathcal{B}v^c = \mathcal{B}w^c$. Furthermore, there exist r_v, $r_w \in L^p(0,T;\mathcal{Q}^*)$ such that $v^c = \mathcal{B}^- r_v$ and $w^c = \mathcal{B}^- r_w$. By Assumption 3.1 we obtain

$$r_v - r_w = \mathcal{B}\mathcal{B}^- r_v - \mathcal{B}\mathcal{B}^- r_w = \mathcal{B}v^c - \mathcal{B}w^c = 0.$$

Thus, it holds that $v^c = \mathcal{B}^- r_v = \mathcal{B}^- r_w = w^c$ and finally also $v_0 = w_0$. □

Lemma 3.3. *Let $\mathcal{W}$ be a closed subspace of $\mathcal{V}$ such that there exists a projection $\mathcal{P}\colon \mathcal{V} \to \mathcal{V}$ that maps $\mathcal{V}$ onto $\mathcal{W}$ and consider $v \in L^p(0,T;\mathcal{W})$. Then, the existence of a time derivative $\dot v \in L^p(0,T;\mathcal{V})$ implies $\dot v \in L^p(0,T;\mathcal{W})$.*

Proof. Assume $v \in L^p(0,T;\mathcal{W})$ with $\dot v \in L^p(0,T;\mathcal{V})$. By assumption, it holds that $(\mathrm{id} - \mathcal{P})v(t) = 0$ for almost all $t \in (0,T)$ with id denoting the identity. Since the time derivative of v exists in a generalized sense [36, Ch. 23.5], we can write $(\mathrm{id} - \mathcal{P})\dot v(t) = 0$, which implies for $t \in (0,T)$ a.e.,

$$\dot v(t) = \mathcal{P}\dot v(t) \in \mathcal{W}.$$ □

3.2. Reformulation. This subsection is devoted to the reformulation and regularization of the operator DAE (3). This extends the results of [4] to a setup with a time-dependent constraint. In Section 3.3 we discuss the resulting positive effects in terms of the sensitivity to perturbations. In Section 4 we then show that the reformulation is in fact an index reduction on operator level and thus, a regularization.

We adapt the technique of minimal extension, cf. Section 2.3, to the operator case. For this, we first add to system (3) the time derivative of the constraint,

$$\mathcal{B}(t)\dot u + \dot{\mathcal{B}}(t)u = \dot{\mathcal{G}}(t).$$

Clearly, this requires the right-hand side $\mathcal{G}$ to be differentiable in the generalized sense, i.e., $\mathcal{G} \in W^{1,p}(0,T;\mathcal{Q}^*)$. Note that this assumption is already needed for the existence of a solution of (3). This fact comes from the theory of DAEs, see for example [20, Th.2.29], which shows that even for the finite-dimensional case with constant coefficients higher derivatives of the right-hand side are necessary. At this point, also $\dot u \in L^p(0,T;\mathcal{V})$ seems to be a necessary condition. However, as the next paragraph shows, this requirement applies only to a part of $\dot u$.

Second, we use the decomposition from Lemma 3.2 to split u into $u_1 \in L^p(0,T;\mathcal{V}_\mathcal{B})$ and $u_2 \in L^p(0,T;\mathcal{V}^c)$. Therewith, the two constraints reduce to

$$\mathcal{B}(t)u_2 = \mathcal{G}(t) \qquad \text{and} \qquad \mathcal{B}(t)\dot u_2 + \dot{\mathcal{B}}(t)u_2 = \dot{\mathcal{G}}(t).$$

Thus, it is sufficient that the derivative of u_2 is an element of $\mathcal{V}$. For u as a whole, we only need that $\dot{u} \in L^q(0,T;\mathcal{V}^*)$. The assumed regularity of $\mathcal{G}$ implies with Assumption 3.1, Lemma 3.1, and equation (3b) that $u_2 \in W^{1,p}(0,T;\mathcal{V}^c)$.

Having added one equation, in a third step, we introduce a new variable $v_2 := \dot{u}_2 \in L^p(0,T;\mathcal{V}^c)$. Recall that $\mathcal{V}^c$ is a subspace of $\mathcal{V}$ for which there exists a projection, cf. Lemma 3.1. Thus, we can apply Lemma 3.3 at this point. The addition of a new variable compensates the redundancy of the two constraints. Note that in the reformulated system the variable u_2 is not differentiated anymore such that we only need an initial condition for u_1. The initial condition for u_2 in the original formulation corresponds to a consistency condition, which typically appears for DAEs [20, Ch. 1]. The overall system then reads: for given data $\mathcal{F} \in L^q(0,T;\mathcal{V}^*)$ and $\mathcal{G} \in W^{1,p}(0,T;\mathcal{Q}^*)$ find functions $u_1 \in L^p(0,T;\mathcal{V}_\mathcal{B})$ with $\dot{u}_1 \in L^q(0,T;\mathcal{V}^*)$, u_2, $v_2 \in L^p(0,T;\mathcal{V}^c)$, and $\lambda \in L^{p'}(0,T;\mathcal{Q})$ such that

$$\dot{u}_1(t) + v_2(t) + \mathcal{K}(t)\big(u_1(t) + u_2(t)\big) + \mathcal{B}^*(t)\lambda(t) = \mathcal{F}(t) \quad \text{in } \mathcal{V}^*, \tag{5a}$$

$$\mathcal{B}(t)u_2(t) = \mathcal{G}(t) \quad \text{in } \mathcal{Q}^*, \tag{5b}$$

$$\mathcal{B}(t)v_2(t) + \dot{\mathcal{B}}(t)u_2(t) = \dot{\mathcal{G}}(t) \quad \text{in } \mathcal{Q}^* \tag{5c}$$

holds for $t \in (0,T)$ a.e. with the initial condition

$$u_1(0) = a_0 := a - \mathcal{B}^-(0)\mathcal{G}(0) \in \mathcal{H}. \tag{5d}$$

The initial condition is well-posed for time smooth $\mathcal{G}$, since $W^{1,p}(0,T;\mathcal{Q}^*)$ is continuously embedded in the space of continuous functions with values in $\mathcal{Q}^*$, namely $C([0,T],\mathcal{Q}^*)$ [28, Lem. 7.1]. Further, note that $L^{p'}(0,T;\mathcal{Q})$ is the right space for the multiplier λ, since for a separable Banach space $\mathcal{Q}^*$ the dual space of $L^{p'}(0,T;\mathcal{Q})$ can be identified with $L^p(0,T;\mathcal{Q}^*)$, cf. [28, Prop. 1.38].

In the following theorem, we discuss the connection of the original system (3) and the regularized formulation (5), cf. [4, Th. 2.3]. In the sequel, we omit to write the time-dependency of the operators $\mathcal{K}$ and $\mathcal{B}$.

Theorem 3.4 (Equivalence of the reformulation)**.** *Consider exponents $1 < q \le p < \infty$, and p' with $1/p + 1/p' = 1$. Assume that $\mathcal{F} \in L^q(0,T;\mathcal{V}^*)$, $\mathcal{G} \in W^{1,p}(0,T;\mathcal{Q}^*)$, and $a \in \mathcal{H}$ as well as the operator $\mathcal{B}$ satisfying Assumption 3.1. Then, the operator DAE (3) has a solution (u,λ) with $u \in L^p(0,T;\mathcal{V})$, $\dot{u} \in L^q(0,T;\mathcal{V}^*)$, and $\lambda \in L^{p'}(0,T;\mathcal{Q})$ if and only if system (5) has a solution (u_1,u_2,v_2,λ) with $u_1 \in L^p(0,T,\mathcal{V}_\mathcal{B})$, $\dot{u}_1 \in L^q(0,T;\mathcal{V}^*)$, u_2, $v_2 \in L^p(0,T;\mathcal{V}^c)$, and $\lambda \in L^{p'}(0,T;\mathcal{Q})$. Furthermore, it holds that $u = u_1 + u_2$ and $\dot{u}_2 = v_2$.*

Proof. Let (u,λ) be a solution of (3). We define

$$u_1 := u - \mathcal{B}^-\mathcal{B}u \in L^p(0,T;\mathcal{V}_\mathcal{B}) \quad \text{and} \quad u_2 := \mathcal{B}^-\mathcal{B}u \in L^p(0,T;\mathcal{V}^c).$$

With equation (3b), we obtain $u_2 = \mathcal{B}^-\mathcal{G}$ and thus, by the regularity of $\mathcal{G}$ and Assumption 3.1, $\dot{u}_2 \in L^p(0,T;\mathcal{V}^c)$. With $v_2 := \dot{u}_2$ the quadruple (u_1,u_2,v_2,λ) satisfies equations (5a-c). The initial condition (5d) is satisfied because of

$$u_1(0) = u(0) - u_2(0) = a - \mathcal{B}^-\mathcal{G}(0).$$

For the reverse direction consider a solution of (5), namely (u_1,u_2,v_2,λ). Then, $u := u_1 + u_2 \in L^p(0,T;\mathcal{V})$. Because of the regularity of $\mathcal{G}$, equation (5b), and $q \le p$, it holds that $\dot{u} = \dot{u}_1 + \dot{u}_2 \in L^q(0,T;\mathcal{V}^*)$. We show that $\dot{u}_2 = v_2$. Equation (5c) and

the time derivative of equation (5b) yield

$$\mathcal{B}v_2 + \dot{\mathcal{B}}u_2 = \dot{\mathcal{G}} = \frac{\mathrm{d}}{\mathrm{d}t}\big(\mathcal{B}u_2\big) = \mathcal{B}\dot{u}_2 + \dot{\mathcal{B}}u_2.$$

Note that $\dot{u}_2 \in L^p(0,T;\mathcal{V}^c)$, as shown in the first part of the proof. The invertibility of $\mathcal{B}$ on $\mathcal{V}^c$, see Lemma 3.1, then gives $\dot{u}_2 = v_2$. Thus, the pair (u,λ) satisfies equations (3a) and (3b). For the initial condition (3c), we obtain

$$u(0) = u_1(0) + u_2(0) = a - \mathcal{B}^-\mathcal{G}(0) + \mathcal{B}^-\mathcal{G}(0) = a. \qquad \square$$

From the solution representation given in Theorem 3.4 we deduce that not every initial condition $a \in \mathcal{H}$ admits a solution to (3).

Corollary 3.5. *Let the assumptions of Theorem 3.4 hold. For the existence of a solution to* (3) *it is necessary that the initial data $a \in \mathcal{H}$ can be decomposed as $a = a_0 + \mathcal{B}^-\mathcal{G}(0)$, where $\mathcal{B}^-\mathcal{G}(0) \in \mathcal{V}^c$ and a_0 is in the closure of $\mathcal{V}_\mathcal{B}$ in $\mathcal{H}$.*

We discuss some examples for which we obtain different kinds of consistency conditions.

Example 3.2. If the operator $\mathcal{B}$ equals the divergence operator and $\mathcal{V} = [H_0^1(\Omega)]^d$, then $\mathcal{V}_\mathcal{B}$ denotes the space of divergence-free functions in $\mathcal{V}$. In this case, the closure of $\mathcal{V}_\mathcal{B}$ with respect to $\mathcal{H} = [L^2(\Omega)]^d$ is a proper subspace of $\mathcal{H}$, cf. [33, Ch. 1, Thm. 1.4],

$$\overline{\mathcal{V}_\mathcal{B}}^{\mathcal{H}} = \big\{v \in \mathcal{H} \mid \nabla\cdot v = 0,\ v\cdot\nu_{\partial\Omega} = 0\big\} \neq \mathcal{H}.$$

Therein, $\nu_{\partial\Omega}$ denotes the normal outer vector along the boundary. Note that the closure is even a subspace of $H(\mathrm{div},\Omega) = \{v \in \mathcal{H} \mid \nabla\cdot v \in L^2(\Omega)\}$. Thus, the initial value a_0 cannot be chosen arbitrarily in $\mathcal{H}$.

Example 3.3. If $\mathcal{B}$ equals the trace operator, i.e., $\mathcal{B}\colon \mathcal{V} := H^1(\Omega) \to H^{1/2}(\partial\Omega)$, then we have $\mathcal{V}_\mathcal{B} = H_0^1(\Omega)$. Since the closure of $H_0^1(\Omega)$ in $\mathcal{H} := L^2(\Omega)$ equals $\mathcal{H}$ itself, the initial data only has to satisfy $a_0 \in \mathcal{H}$. In fact, this means that the initial data $a = a_0 + \mathcal{B}^-\mathcal{G}(0)$ can also be chosen arbitrarily in $\mathcal{H}$ such that there is no consistency condition in this case.

3.3. Influence of perturbations. As mentioned in the introduction, we do not define an index for operator DAEs. Nevertheless, the influence of perturbations provides information about the stability of the system, similar to the *perturbation index* for DAEs. We show the positive effect of the presented regularization in terms of perturbations. For this, we restrict the analysis to the case $p = q = 2$ with a linear, symmetric, on $\mathcal{V}_\mathcal{B}$ uniformly elliptic, and uniformly bounded operator $\mathcal{K}$, i.e, for $u \in \mathcal{V}_\mathcal{B}$ and v, $w \in \mathcal{V}$ we assume that

$$\langle \mathcal{K}(t)u, u\rangle \geq k_1\|u\|^2 \quad \text{and} \quad \langle \mathcal{K}(t)v, w\rangle \leq k_2\|v\|\|w\|.$$

Note that we use $\|\cdot\| := \|\cdot\|_\mathcal{V}$ and later $|\cdot| := \|\cdot\|_\mathcal{H}$ to simplify the notation. We consider the to (3) corresponding perturbed problem

$$\dot{\bar{u}} + \mathcal{K}\bar{u} + \mathcal{B}^*\bar{\lambda} = \mathcal{F} + \delta \quad \text{in } \mathcal{V}^*, \tag{6a}$$

$$\mathcal{B}\bar{u} = \mathcal{G} + \theta \quad \text{in } \mathcal{Q}^*. \tag{6b}$$

Here, $(\bar{u}, \bar{\lambda})$ denotes the solution if we include perturbations $\delta\colon [0,T] \to \mathcal{V}^*$ and $\theta\colon [0,T] \to \mathcal{Q}^*$. For the regularized equations, the perturbed problem has the form

$$\dot{\hat{u}}_1 + \hat{v}_2 + \mathcal{K}(\hat{u}_1 + \hat{u}_2) + \mathcal{B}^*\hat{\lambda} = \mathcal{F} + \delta \quad \text{in } \mathcal{V}^*, \tag{7a}$$

$$\mathcal{B}\hat{u}_2 = \mathcal{G} + \theta \quad \text{in } \mathcal{Q}^*, \tag{7b}$$

$$\mathcal{B}\hat{v}_2 + \dot{\mathcal{B}}\hat{u}_2 = \dot{\mathcal{G}} + \xi \quad \text{in } \mathcal{Q}^*. \tag{7c}$$

For this system, we consider perturbations of the form $\delta \in L^2(0,T;\mathcal{V}^*)$ and θ, $\xi \in L^2(0,T;\mathcal{Q}^*)$ and the solution is denoted by $(\hat{u}_1, \hat{u}_2, \hat{v}_2, \hat{\lambda})$. The initial condition is given by $\hat{u}_1(0) = u_1(0) - e_{1,0}$, i.e., $e_{1,0}$ contains the initial error. Because of Theorem 3.4, it is sufficient to consider the regularized system (7). The result for the original operator DAE then follows if we replace ξ by $\dot{\theta}$, cf. Remark 3.4 below.

By C_{emb} we denote the continuity constant of the embedding $\mathcal{V} \hookrightarrow \mathcal{H}$. Furthermore, we introduce the errors

$$e_1 := \hat{u}_1 - u_1, \qquad e_2 := \hat{u}_2 - u_2, \qquad e_v := \hat{v}_2 - v_2, \qquad e_\lambda := \hat{\lambda} - \lambda.$$

Theorem 3.6. *Consider the perturbed problem* (7) *with a linear, symmetric, on $\mathcal{V}_\mathcal{B}$ uniformly elliptic, and uniformly bounded operator $\mathcal{K}$. Furthermore, let $\mathcal{B}$ satisfy Assumption 3.1 and the perturbations $\delta \in L^2(0,T;\mathcal{V}^*)$ and θ, $\xi \in L^2(0,T;\mathcal{Q}^*)$. Then, the error in the differential variable u_1 satisfies with a positive constant $c \in \mathbb{R}$ that*

$$\begin{aligned}&\|e_1\|^2_{C([0,T];\mathcal{H})} + k_1\|e_1\|^2_{L^2(0,T;\mathcal{V})}\\ \le& |e_{1,0}|^2 + c\Big[\|\delta\|^2_{L^2(0,T;\mathcal{V}^*)} + \|\theta\|^2_{L^2(0,T;\mathcal{Q}^*)} + \|\xi\|^2_{L^2(0,T;\mathcal{Q}^*)}\Big].\end{aligned}$$

Proof. Bounds of the errors e_2 and e_v can be easily found by the continuity of $\dot{\mathcal{B}}$ and the right-inverse of $\mathcal{B}$. For the error in the differential part u_1, we test the difference of equations (7a) and (5a) by $e_1 \in \mathcal{V}_\mathcal{B}$. Thus, the term with the Lagrange multiplier vanishes and the properties of the operator $\mathcal{K}$ can be exploited. The details can be found in [3, Sect. 6.1]. □

Remark 3.4. In order to transfer these results to the perturbation analysis of the original formulation (3) we have to insert $\xi = \dot{\theta}$. Thus, the error also depends on the derivative of the perturbation θ. This leads to possible instabilities known from high-index DAEs. If the perturbation is not smooth, it may even make the model useless. In this regard, the presented reformulation can be seen as a regularization of the system.

In the given setting of the evolution equations, it is not possible to gain similar estimates for e_λ. Estimates of the error in the Lagrange multiplier are only possible if we consider the primitive of e_λ or assume more regular perturbations $\delta \in L^2(0,T;\mathcal{H}^*)$ and $e_{1,0} \in \mathcal{V}_\mathcal{B}$.

4. Spatial discretization

For the simulation of time-dependent PDEs, we need discretizations in time and space. Because of the special role of the time variable in DAEs we only consider the approach of discretizing in space and time separately. In this section, we consider the systems, which result from a spatial discretization of the operator DAE. Thus, we follow the *method of lines*. The *Rothe method* [27], in which one discretizes in time first, is then discussed in Section 5.

As mentioned above, a spatial discretization of an operator DAE leads to a classical DAE, for which the differentiation index is well-defined [8, 16, 20]. Within this section, we simply write *index*, meaning the differentiation index, cf. Section 2.3. Recall that we do not use any index definition for PDAEs.

We show that the DAE corresponding to the original system (3) is of index 2 whereas the DAEs resulting from the reformulated systems are of index 1. For this, only standard assumptions on the used finite element schemes have to be considered. This then shows that also in this sense the reformulation presented in Section 3.2 is a regularization.

4.1. Finite element discretization. For the spatial discretization, we consider finite-dimensional approximations of the spaces $\mathcal{V}_{\mathcal{B}}$, $\mathcal{V}^{\mathrm{c}}$, and $\mathcal{Q}$. We denote the approximation spaces by $V_{\mathcal{B},h}$, V_h^{c}, and Q_h, respectively. Furthermore, we define $V_h = V_{\mathcal{B},h} \oplus V_h^{\mathrm{c}}$ as finite-dimensional approximation of $\mathcal{V}$.

Thinking of finite elements on a regular mesh $\mathcal{T}$ of the domain Ω, cf. [7], we consider basis functions $\{\varphi_i\}_{1,\dots,n_1}$ of $V_{\mathcal{B},h}$, $\{\varphi_i\}_{n_1+1,\dots,n}$ of V_h^{c}, and $\{\psi_i\}_{1,\dots,m}$ of Q_h with $m = n - n_1$. Hence, we assume that $\dim V_h^{\mathrm{c}} = \dim Q_h$. The finite-dimensional approximations of u_1, u_2, v_2, and λ are then represented by the coefficient vectors q_1, q_2, r_2, and μ, respectively. By $q \in \mathbb{R}^n$ we denote the vector $q = [q_1^T, q_2^T]^T$. Based on this discretization scheme, we define the positive definite mass matrix $M \in \mathbb{R}^{n,n}$ by $M_{i,j} := (\varphi_i, \varphi_j)$. The discrete version of the constraint operator $\mathcal{B}$ is defined by

$$B(t) \in \mathbb{R}^{m,n}, \qquad B_{j,i}(t) := \big\langle \mathcal{B}(t)\varphi_i, \psi_j \big\rangle. \tag{8}$$

Note that, according to Assumption 3.1, it is natural to assume that B is continuously differentiable with respect to time and that B has full rank.

Remark 4.1 (Nonconforming discretization). In order that B is well-defined, the operator $\mathcal{B}$ has to be defined for the given basis functions. Since nonconforming finite elements are not excluded [7, Ch. III], the application of $\mathcal{B}$ may be generalized to an elementwise application.

Remark 4.2 (Inf-sup stability). For the unique solvability of the semi-discrete systems resulting from the finite element discretization it is sufficient that the constraint matrix B is of full rank. For a stable approximation of the Lagrange multiplier λ in terms of the discretization parameter h, one may assume an inf-sup condition, i.e., there exists a constant $\beta_{\mathrm{disc}} > 0$, independent of h and time t, such that

$$\inf_{q_h \in Q_h} \sup_{v_h \in V_h} \frac{\langle \mathcal{B}(t)v_h, q_h \rangle}{\|v_h\| \, \|q_h\|_{\mathcal{Q}}} \geq \beta_{\mathrm{disc}},$$

cf., e.g., [7, Ch. III.4].

Finally, we denote the discrete version of the possibly time-dependent operator $\mathcal{K}$ by $K(t)\colon \mathbb{R}^n \to \mathbb{R}^n$, which may be written as a time-dependent $n \times n$ matrix in the linear case.

4.2. Discretization of (3). For the computation of the index of the resulting DAE system, we assume that the mass matrix M is positive definite and that B is of full rank. First, we consider the index of the DAE, which results from a spatial discretization of the original system (3). With the introduced notation, the DAE

has the form

$$M\dot{q} + K(t,q) + B^T(t)\mu = f, \tag{9a}$$

$$B(t)q = g. \tag{9b}$$

From Lemma 2.3 we infer that the system (9) is of index 2, since $BM^{-1}B^T$ is invertible for all times t. The index-2 structure can also made visible through a differentiation of the constraint (9b). This then leads to the (analytically) equivalent DAE

$$\begin{bmatrix} M & B^T(t) \\ B(t) & 0 \end{bmatrix} \begin{bmatrix} \dot{q} \\ \mu \end{bmatrix} = \begin{bmatrix} f - K(t,q) \\ \dot{g} - \dot{B}(t)q \end{bmatrix}.$$

Again the assumptions on M and B imply that the matrix on the left-hand side is invertible. Thus, a single differentiation leads to an ODE for q and an algebraic equation for μ.

4.3. Discretization of (5)**.** In the case of a conforming discretization, i.e., $V_{\mathcal{B},h} \subseteq \mathcal{V}_{\mathcal{B}}$, $V_h^{\mathrm{c}} \subseteq \mathcal{V}^{\mathrm{c}}$, and $Q_h \subseteq \mathcal{Q}$, the matrix $B(t)$ has the special structure $B(t) = [0 \;\; B_2(t)]$. Therein, the matrix $B_2(t)$ is square and non-singular. In this case, the semi-discrete version of (5) has the form

$$M \begin{bmatrix} \dot{q}_1 \\ r_2 \end{bmatrix} + K(t,q_1,q_2) + \begin{bmatrix} 0 \\ B_2^T(t) \end{bmatrix} \mu = f, \tag{10a}$$

$$B_2(t)q_2 = g, \tag{10b}$$

$$B_2(t)r_2 = \dot{g} - \dot{B}_2(t)q_2. \tag{10c}$$

This system forms a DAE of index 1 as we show in Lemma 4.1 below. In many cases, one depends on a nonconforming spatial discretization [7, Ch. III], i.e., the discrete ansatz spaces are not subspaces of the original search spaces. One simple example is the Crouzeix-Raviart element [10], a lowest order piecewise linear but discontinuous discretization scheme. Since we do not assume $V_{\mathcal{B},h} \subseteq \mathcal{V}_{\mathcal{B}}$ for general mixed finite element discretizations, i.e., $\ker B(t) \not\subseteq \ker \mathcal{B}(t)$, cf. [15, Ch. 3], we lose the special structure of $B(t)$.

In general, we have $B(t) = [B_1(t) \; B_2(t)]$ and simply assume that the block B_2 is non-singular. This is no restriction, since one may always permute the columns of B (corresponds to a reordering of basis functions in $V_{\mathcal{B},h}$ and V_h^{c}) such that the B_2 block is regular. Then, the semi-discretized system reads

$$M \begin{bmatrix} \dot{q}_1 \\ r_2 \end{bmatrix} + K(t,q_1,q_2) + B^T(t)\mu = f, \tag{11a}$$

$$B_2(t)q_2 = g - B_1(t)q_1, \tag{11b}$$

$$B_2(t)r_2 = \dot{g} - B_1(t)\dot{q}_1 - \dot{B}_1(t)q_1 - \dot{B}_2(t)q_2. \tag{11c}$$

Lemma 4.1 (Index-1 DAE)**.** *For a positive definite mass matrix M and a differentiable constraint matrix B with a regular block B_2, the DAEs* (10) *and* (11) *are of index 1.*

Proof. Similar to the proof of [20, Th. 6.12], we show that (11) is of index 1. The property then follows for system (10) as well because it is a special case.

Since the matrix $B_2(t)$ is of full rank, equations (11b) and (11c) yield direct expressions of q_2 and r_2 in terms of q_1 and $\dot{q}_1$. Furthermore, a multiplication of (11a) from the left by BM^{-1} provides a formula for μ in terms of q_1. Here we use

the assumptions on M and B, which imply that the matrix $BM^{-1}B^T$ is invertible. Finally, inserting all these expressions into equation (11a), we obtain an ODE in q_1. Thus, we can solve system (11) without any further differentiation steps. □

4.4. Commutativity. The connection between the presented regularization on operator level in Section 3 and the index reduction for DAEs is illustrated in Figure 4.1, cf. [2] for second-order systems. The scheme shows that regularization and spatial discretization commute if corresponding discretization schemes are used and the index reduction is performed by minimal extension, cf. Section 2.3. Note that this is beneficial for adaptive simulations as modifications of the finite element schemes such as a refinement of the underlying mesh do not call for another index reduction step afterwards. This fact is also of importance for the Rothe discretization as shown in the numerical example in Section 6.

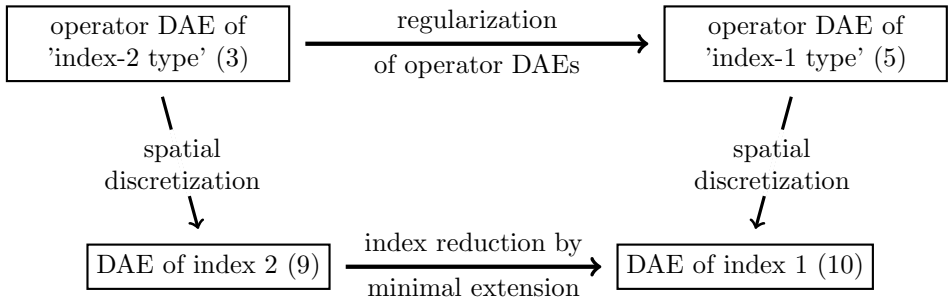

FIGURE 4.1. Illustrative scheme of the commutativity of regularization and spatial discretization.

Since the index of the DAE (10) is, compared to the DAE (9), reduced by one, we may call the proposed regularization procedure from Section 3 an index reduction on operator level.

5. Temporal discretization

For the time integration of the operator DAE (5) we restrict ourselves to the implicit Euler method. We prove the convergence of the scheme and highlight the needed adjustments in contrast to operator ODEs (time-dependent PDEs without constraint), for which the convergence is well-known. Again we benefit from the regularization introduced in Section 3 because of the obtained robustness with respect to perturbations. Furthermore, we consider perturbations of the right-hand sides in order to analyse the convergence of the Rothe method applied to operator DAEs.

5.1. Time-discrete systems. As in Section 3.3, we restrict the analysis to the linear case with $p = q = 2$. Furthermore, we restrict the analysis to the time-independent case, i.e., $\mathcal{K}(t) \equiv \mathcal{K}$ and $\mathcal{B}(t) \equiv \mathcal{B}$. This means that we assume $\mathcal{K}$ to be linear, symmetric, continuous, and positive on $\mathcal{V}_\mathcal{B}$. Furthermore, we assume $\mathcal{B}$ to have a time-independent right-inverse $\mathcal{B}^-\colon \mathcal{Q}^* \to \mathcal{V}^c$ with continuity constant $C_{\mathcal{B}^-}$, cf. Assumption 3.1. Recall that $(\cdot,\cdot) := (\cdot,\cdot)_\mathcal{H}$ denotes the inner product in the space $\mathcal{H}$.

For the temporal discretization we consider an equidistant partition $0 = t_0 < t_1 < \cdots < t_n = T$ of the interval $[0,T]$ with time step size τ. The semi-discrete approximations of u_1, u_2, v_2, and λ at time $t_j = j\tau$ are denoted by u_1^j, u_2^j, v_2^j, and λ^j, respectively. For the application of the implicit Euler scheme to system (5) we replace the derivative $\dot{u}_1$ by the *discrete derivative* $Du_1^j := (u_1^j - u_1^{j-1})/\tau$. This then leads to a sequence of stationary PDEs, which have to be solved in every time step. The differential equation (5a) turns into

$$(Du_1^j, v) + \langle \mathcal{K}u_1^j, v\rangle + \langle \lambda^j, \mathcal{B}v\rangle = \langle \mathcal{F}^j, v\rangle - (v_2^j, v) - \langle \mathcal{K}u_2^j, v\rangle \tag{12}$$

for $j = 1, \dots, n$, whereas the constraints (5b) and (5c) result in

$$\langle \mathcal{B}u_2^j, q\rangle = \langle \mathcal{G}^j, q\rangle, \qquad \langle \mathcal{B}v_2^j, q\rangle = \langle \dot{\mathcal{G}}^j, q\rangle. \tag{13}$$

Therein, we assume that $u_1^{j-1} \in \mathcal{H}$ is given and search for $u_1^j \in \mathcal{V}_\mathcal{B}$, $u_2^j, v_2^j \in \mathcal{V}^\mathrm{c}$, and $\lambda^j \in \mathcal{Q}$. The test functions are given by $v \in \mathcal{V}$ and $q \in \mathcal{Q}$.

Since $\mathcal{G}$ is continuous, we set $\mathcal{G}^j = \mathcal{G}(t_j)$. Note, however, that $\mathcal{F}^j$ and $\dot{\mathcal{G}}^j$ cannot equal the function evaluations of $\mathcal{F}$ and $\dot{\mathcal{G}}$ at time t_j, since this is not well-defined. Instead, we use integral means over a time interval, i.e.,

$$\mathcal{F}^j := \frac{1}{\tau}\int_{t_{j-1}}^{t_j} \mathcal{F}(t)\,\mathrm{d}t \in \mathcal{V}^*, \qquad \dot{\mathcal{G}}^j := \frac{1}{\tau}\int_{t_{j-1}}^{t_j} \dot{\mathcal{G}}(t)\,\mathrm{d}t \in \mathcal{Q}^*.$$

With these approximations we define $\mathcal{F}_\tau$, $\mathcal{G}_\tau$, and $\dot{\mathcal{G}}_\tau$ as the piecewise constants

$$\mathcal{F}_\tau(t) := \mathcal{F}^j, \qquad \mathcal{G}_\tau(t) := \mathcal{G}^j, \qquad \dot{\mathcal{G}}_\tau(t) := \dot{\mathcal{G}}^j,$$

for $t \in]t_j, t_{j+1}]$ and continuous extension in $t = 0$. This then implies

$$\mathcal{F}_\tau \to \mathcal{F} \text{ in } L^2(0,T;\mathcal{V}^*), \qquad \mathcal{G}_\tau \to \mathcal{G},\ \dot{\mathcal{G}}_\tau \to \dot{\mathcal{G}} \text{ in } L^2(0,T;\mathcal{Q}^*). \tag{14}$$

Since the operator $\mathcal{B}$ is invertible on $\mathcal{V}^\mathrm{c}$, we argue from (13) that $u_2^j = \mathcal{B}^-\mathcal{G}^j$ and $v_2^j = \mathcal{B}^-\dot{\mathcal{G}}^j$. Inserting this into (12) and testing only with functions $v \in \mathcal{V}_\mathcal{B}$, we obtain

$$(Du_1^j, v) + \langle \mathcal{K}u_1^j, v\rangle = \langle \mathcal{F}^j, v\rangle - (\mathcal{B}^-\dot{\mathcal{G}}^j, v) - \langle \mathcal{K}\mathcal{B}^-\mathcal{G}^j, v\rangle. \tag{15}$$

This equation will be used for the stability estimates in the following subsection. It remains to discuss the solvability of system (12) for u_1^j and λ^j. With the bilinear form $c\colon \mathcal{V}\times\mathcal{V} \to \mathbb{R}$, given by $c(u,v) := \tau^{-1}(u,v) + \langle \mathcal{K}u, v\rangle$, and the functional $F \in \mathcal{V}^*$,

$$\langle F, v\rangle := \langle \mathcal{F}^j, v\rangle - (\mathcal{B}^-\dot{\mathcal{G}}^j, v) - \langle \mathcal{K}\mathcal{B}^-\mathcal{G}^j, v\rangle + \frac{1}{\tau}(u_1^{j-1}, v),$$

equation (15) can be written as $c(u_1^j, v) = F(v)$ for all $v \in \mathcal{V}_\mathcal{B}$. The unique solvability of (15) for u_1^j then follows by the Lax-Milgram lemma [14, Sect. 6.2.1]. For the unique solvability of λ^j we consider equation (12) tested by functions $v \in \mathcal{V}^\mathrm{c}$,

$$\langle \mathcal{B}^*\lambda^j, v\rangle = \langle \mathcal{F}^j, v\rangle - (\mathcal{B}^-\dot{\mathcal{G}}^j, v) - \langle \mathcal{K}\mathcal{B}^-\mathcal{G}^j, v\rangle - (Du_1^j, v) - \langle \mathcal{K}u_1^j, v\rangle. \tag{16}$$

Equation (15) implies that the right-hand side of (16) vanishes for all functions in $\mathcal{V}_\mathcal{B}$. Thus, the functional is an element of the polar set $\mathcal{V}_\mathcal{B}^\mathrm{o}$, on which the operator $\mathcal{B}^*$ is invertible [7, Ch. III, Lem. 4.2].

5.2. Stability estimates. The most important ingredient for the convergence analysis of the Rothe method are stability or a priori estimates, which we provide in this subsection. Since equation (15) is essentially an operator ODE for u_1^j, the shown bounds follow the lines of the stability results in [11, Ch. 4], see also [33, Ch. III.4]. Amongst others, we take advantage of the equality

$$2(Du^j, u^j) = D|u^j|^2 + \tau|Du^j|^2. \tag{17}$$

This identity follows by a simple calculation, cf. [11, Lem. 3.2.2]. For the differential variable u_1^j we obtain the following result.

Lemma 5.1 (Stability)**.** *Assume $\mathcal{F} \in L^2(0,T;\mathcal{V}_{\mathcal{B}}^*)$ and $\mathcal{G} \in H^1(0,T;\mathcal{Q}^*)$. Then, the approximations $u_1^k \in \mathcal{V}_{\mathcal{B}}$ given by the Euler scheme* (15) *with $u_1^0 \in \mathcal{H}$ satisfy for all $1 \le k \le n$ the estimate*

$$|u_1^k|^2 + \tau^2 \sum_{j=1}^{k} |Du_1^j|^2 + \tau k_1 \sum_{j=1}^{k} \|u_1^j\|^2 \le M^2 \tag{18}$$

with the constant $M^2 := |u_1^0|^2 + 3k_1^{-1}\big(\|\mathcal{F}\|^2_{L^2(0,T;\mathcal{V}_{\mathcal{B}}^*)} + C^2_{\mathcal{B}^-}(C^4_{emb}+k_2^2)\|\mathcal{G}\|^2_{H^1(0,T;\mathcal{Q}^*)}\big)$.

Proof. Using as test function $v = u_1^j \in \mathcal{V}_{\mathcal{B}}$, $j \ge 1$, in the implicit Euler scheme (15), we obtain

$$(Du_1^j, u_1^j) + \langle \mathcal{K}u_1^j, u_1^j\rangle = \langle \mathcal{F}^j, u_1^j\rangle - (\mathcal{B}^-\dot{\mathcal{G}}^j, u_1^j) - \langle \mathcal{K}\mathcal{B}^-\mathcal{G}^j, u_1^j\rangle. \tag{19}$$

Summation over $j = 1, \dots, k$, together with property (17), the Cauchy-Schwarz inequality, and the continuous embedding $\mathcal{V} \hookrightarrow \mathcal{H}$ yields

$$\begin{aligned}
|u_1^k|^2 - |u_1^0|^2 + \tau^2 \sum_{j=1}^{k} |Du_1^j|^2 + 2\tau k_1 \sum_{j=1}^{k} \|u_1^j\|^2 &\\
&\overset{(17)}{\le} 2\tau \sum_{j=1}^{k} (Du_1^j, u_1^j) + 2\tau \sum_{j=1}^{k} \langle \mathcal{K}u_1^j, u_1^j\rangle \\
&\overset{(19)}{\le} 2\tau \sum_{j=1}^{k} \Big(\|\mathcal{F}^j\|_{\mathcal{V}_{\mathcal{B}}^*} + C_{\mathrm{emb}}|\mathcal{B}^-\dot{\mathcal{G}}^j| + k_2\|\mathcal{B}^-\mathcal{G}^j\|\Big)\|u_1^j\|.
\end{aligned}$$

The application of Young's inequality [14, App. B] and the boundedness of $\mathcal{B}^-$ shows that

$$\begin{aligned}
|u_1^k|^2 - |u_1^0|^2 + \tau^2 \sum_{j=1}^{k} |Du_1^j|^2 + \tau k_1 \sum_{j=1}^{k} \|u_1^j\|^2 &\\
&\le \frac{3\tau}{k_1} \sum_{j=1}^{k} \Big(\|\mathcal{F}^j\|^2_{\mathcal{V}_{\mathcal{B}}^*} + C^2_{\mathcal{B}^-} C^4_{\mathrm{emb}} \|\dot{\mathcal{G}}^j\|^2_{\mathcal{Q}^*} + C^2_{\mathcal{B}^-} k_2^2 \|\mathcal{G}^j\|^2_{\mathcal{Q}^*}\Big).
\end{aligned}$$

Finally, the assertion follows by properties of the Bochner integral, which imply that $\tau \sum_{j=1}^{k} \|\mathcal{F}^j\|^2_{\mathcal{V}_{\mathcal{B}}^*} \le \|\mathcal{F}\|^2_{L^2(0,T;\mathcal{V}_{\mathcal{B}}^*)}$. □

With the same assumptions as in Lemma 5.1, we may also prove that there exists a positive constant $c \in \mathbb{R}$ such that

$$\tau \sum_{j=1}^{n} \|Du_1^j\|^2_{\mathcal{V}_{\mathcal{B}}^*} \le cM^2. \tag{20}$$

This result follows from equation (15), which yields for $j \geq 1$,

$$\begin{aligned}\|Du_1^j\|_{\mathcal{V}_\mathcal{B}^*} &:= \sup_{v\in\mathcal{V}_\mathcal{B},\ \|v\|=1} \left|\langle \mathcal{F}^j, v\rangle - (\mathcal{B}^-\dot{\mathcal{G}}^j, v) - \langle \mathcal{K}\mathcal{B}^-\mathcal{G}^j, v\rangle - \langle \mathcal{K}u_1^j, v\rangle\right| \\ &\leq \ \|\mathcal{F}^j\|_{\mathcal{V}_\mathcal{B}^*} + C_{\mathcal{B}^-}C_{\mathrm{emb}}^2\|\dot{\mathcal{G}}^j\|_{\mathcal{Q}^*} + k_2C_{\mathcal{B}^-}\|\mathcal{G}^j\|_{\mathcal{Q}^*} + k_2\|u_1^j\|.\end{aligned}$$

An application of Young's inequality, the summation over j, and the estimate (18) then finally imply (20).

Remark 5.1. Assume that $(\cdot,\cdot) + \langle \mathcal{K}\cdot,\cdot\rangle$ defines an inner product in $\mathcal{V}$ with respect to which the decomposition $\mathcal{V} = \mathcal{V}_\mathcal{B}\oplus\mathcal{V}^\mathrm{c}$ is orthogonal. If we assume more regularity of the given data in the form of $\mathcal{F} \in L^2(0,T;\mathcal{H}^*)$ and $u_1^0 \in \mathcal{V}_\mathcal{B}$, then we obtain the estimate

$$\tau\sum_{j=1}^n |Du_1^j|^2 \leq M_{\mathrm{reg}}^2.$$

Here, M_{reg} equals the constant M from Lemma 5.1 but with the stronger norms $\|u_1^0\|$ and $\|\mathcal{F}\|_{L^2(0,T;\mathcal{H}^*)}$. To obtain this, equation (15) has to be tested by $Du_1^j \in \mathcal{V}_\mathcal{B}$. Thus, we obtain the estimate (20) in a stronger norm, which is crucial in view of the Lagrange multiplier.

In order to prove the convergence of the Euler scheme, we need to define global approximations. Given u_1^j, $j = 0,\dots,n$, we define the piecewise constant and piecewise linear functions $U_{1,\tau}$ and $\hat{U}_{1,\tau}$ on the interval $[0,T]$ by $U_{1,\tau}(0) = \hat{U}_{1,\tau}(0) = u_1^0$ and for $t \in]t_{j-1}, t_j]$ by

$$U_{1,\tau}(t) := u_1^j, \qquad \hat{U}_{1,\tau}(t) := u_1^j + (t-t_j)Du_1^j. \tag{21}$$

We show that the sequences $U_{1,\tau}$ and $\hat{U}_{1,\tau}$ are uniformly bounded. The boundedness in $L^\infty(0,T;\mathcal{H})$ and $L^2(0,T;\mathcal{V}_\mathcal{B})$ follows directly from (18) if we assume $u_1^0 \in \mathcal{V}_\mathcal{B}$. Additionally, we obtain by the stability estimate (20) that

$$\big\|\dot{\hat{U}}_{1,\tau}\big\|^2_{L^2(0,T;\mathcal{V}_\mathcal{B}^*)} = \sum_{j=1}^n \int_{t_{j-1}}^{t_j} \|Du_1^j\|^2_{\mathcal{V}_\mathcal{B}^*}\,\mathrm{d}t = \tau\sum_{j=1}^n \|Du_1^j\|^2_{\mathcal{V}_\mathcal{B}^*} \leq cM^2. \tag{22}$$

For the approximations of u_2 and v_2 we define similarly the piecewise constant functions

$$U_{2,\tau}(t) := u_2^j \ \text{ if } t \in]t_{j-1}, t_j], \qquad V_{2,\tau}(t) := v_2^j \ \text{ if } t \in]t_{j-1}, t_j] \tag{23}$$

with a continuous extension in $t = 0$. Note that this definition implies that $U_{2,\tau} = \mathcal{B}^-\mathcal{G}_\tau$ and $V_{2,\tau} = \mathcal{B}^-\dot{\mathcal{G}}_\tau$. Finally, we define as approximation of the Lagrange multiplier $\Lambda_\tau\colon]0,T] \to \mathcal{Q}$ by

$$\Lambda_\tau(t) := \lambda^j \ \text{ if } t \in]t_{j-1}, t_j]. \tag{24}$$

With this, we can rewrite equation (16) in the form

$$\langle \mathcal{B}^*\Lambda_\tau, v\rangle = \langle \mathcal{F}_\tau, v\rangle - (\mathcal{B}^-\dot{\mathcal{G}}_\tau, v) - \langle \mathcal{K}\mathcal{B}^-\mathcal{G}_\tau, v\rangle - (\dot{\hat{U}}_{1,\tau}, v) - \langle \mathcal{K}U_{1,\tau}, v\rangle. \tag{25}$$

Here, we consider test functions $v \in \mathcal{V}$ and $t \in (0,T)$ a.e..

Since the Lagrange multiplier corresponds to the algebraic variable in the finite-dimensional case, we expect less regularity than for the other variables. Indeed, we are not able to bound Λ_τ in $L^2(0,T;\mathcal{Q})$ uniformly within the given weak setting. Note that with the additional regularity assumptions from Remark 5.1, we would be able to show the desired boundedness. Instead, we consider the primitive of

Λ_τ, which then leads to a weaker notion of solvability, cf. [13]. We define $\tilde{\Lambda}_\tau \in AC([0,T];\mathcal{Q})$ by

$$\tilde{\Lambda}_\tau(t) := \int_0^t \Lambda_\tau(s)\,\mathrm{d}s. \tag{26}$$

Integrating (25) over $[0,t]$, we obtain an equation for the primitive of the Lagrange multiplier,

$$\langle \mathcal{B}^*\tilde{\Lambda}_\tau, v\rangle = \langle \tilde{\mathcal{F}}_\tau, v\rangle - (\mathcal{B}^-\tilde{\dot{\mathcal{G}}}_\tau, v) - \langle \mathcal{K}\mathcal{B}^-\tilde{\mathcal{G}}_\tau, v\rangle - (\hat{U}_{1,\tau}, v) - \langle \mathcal{K}\tilde{U}_{1,\tau}, v\rangle + (u_1^0, v). \tag{27}$$

Therein, $\tilde{\mathcal{F}}_\tau$, $\tilde{\mathcal{G}}_\tau$, $\tilde{\dot{\mathcal{G}}}_\tau$, and $\tilde{U}_{1,\tau}$ denote the primitives of $\mathcal{F}_\tau$, $\mathcal{G}_\tau$, $\dot{\mathcal{G}}_\tau$, and $U_{1,\tau}$, respectively. For $\tilde{\Lambda}_\tau$ we are able to prove the boundedness independently of the step size τ.

Lemma 5.2 (Boundedness of $\tilde{\Lambda}_\tau$)**.** *Assume $\mathcal{F} \in L^2(0,T;\mathcal{V}^*)$, $\mathcal{G} \in H^1(0,T;\mathcal{Q}^*)$, and $u_1^0 \in \mathcal{V}_\mathcal{B}$. Then, the sequence $\tilde{\Lambda}_\tau$ is bounded in $C([0,T];\mathcal{Q})$.*

Proof. We make use of the inf-sup condition of the operator $\mathcal{B}$ and, by equation (27), we obtain the estimate

$$\begin{aligned}\beta_{\inf}\|\tilde{\Lambda}_\tau\|_{C([0,T];\mathcal{Q})} &\le \max_{t\in[0,T]}\sup_{v\in\mathcal{V}}\frac{\langle \mathcal{B}^*\tilde{\Lambda}_\tau(t), v\rangle}{\|v\|}\\ &\overset{(27)}{\le} \max_{t\in[0,T]}\Big[\|\tilde{\mathcal{F}}_\tau(t)\|_{\mathcal{V}^*} + C_{\mathrm{emb}}|\mathcal{B}^-\tilde{\dot{\mathcal{G}}}_\tau(t)| + k_2\|\mathcal{B}^-\tilde{\mathcal{G}}_\tau(t)\|\\ &\qquad + C_{\mathrm{emb}}|\hat{U}_{1,\tau}(t)| + k_2\|\tilde{U}_{1,\tau}(t)\| + C_{\mathrm{emb}}|u_1^0|\Big].\end{aligned}$$

By the properties of the Bochner integral and (18), this is bounded uniformly in terms of T, the initial data, and the right-hand sides. For details we refer to [3, Sect. 10.3]. □

5.3. Passing to the limit. Since every bounded sequence has a weakly convergent subsequence, the results from the previous subsection imply the existence of weak limits U_1, U_2, V_2, and $\tilde{\Lambda}$. More precisely, we obtain by (14) that

$$U_{2,\tau} = \mathcal{B}^-\mathcal{G}_\tau \to U_2 := \mathcal{B}^-\mathcal{G}, \qquad V_{2,\tau} = \mathcal{B}^-\dot{\mathcal{G}}_\tau \to V_2 := \mathcal{B}^-\dot{\mathcal{G}} \quad \text{in } L^2(0,T;\mathcal{V}^c).$$

The embedding $H^1(0,T;\mathcal{Q}^*) \hookrightarrow C([0,T];\mathcal{Q}^*)$ implies additionally that U_2 satisfies the consistency condition $U_2(0) = \mathcal{B}^-\mathcal{G}(0)$. By the boundedness of $U_{1,\tau}$ and $\hat{U}_{1,\tau}$, we obtain weak limits in $L^2(0,T;\mathcal{V}_\mathcal{B})$. Furthermore, the estimate (18) implies that

$$\big\|U_{1,\tau} - \hat{U}_{1,\tau}\big\|^2_{L^2(0,T;\mathcal{H})} \le \tau\sum_{j=1}^{n}\big|u_1^j - u_1^{j-1}\big|^2 \le \tau M^2 \to 0.$$

Thus, the two limits coincide in $L^2(0,T;\mathcal{H})$ and the continuous embedding $\mathcal{V} \hookrightarrow \mathcal{H}$ implies that the same is true for the limit in $L^2(0,T;\mathcal{V}_\mathcal{B})$. We denote the joined limit by U_1, i.e., $U_{1,\tau}$, $\hat{U}_{1,\tau} \rightharpoonup U_1$ in $L^2(0,T;\mathcal{V}_\mathcal{B})$.

Finally, Lemma 5.2 implies the existence of a weak limit $\tilde{\Lambda}$ for which

$$\tilde{\Lambda}_\tau \rightharpoonup \tilde{\Lambda} \quad \text{in } L^p(0,T;\mathcal{Q})$$

for all $1 < p < \infty$. It remains to show that the obtained limits solve the operator DAE (5). For this, we assume $u_1^0 = a_0 \in \mathcal{V}_\mathcal{B}$. Obviously, the limits U_2 and V_2 solve equations (5b) and (5c). For U_1 we obtain the following result.

Theorem 5.3. *Assume $\mathcal{F} \in L^2(0,T;\mathcal{V}_\mathcal{B}^*)$, $\mathcal{G} \in H^1(0,T;\mathcal{Q}^*)$, and $u_1^0 = a_0 \in \mathcal{V}_\mathcal{B}$. Then, the weak limit $U_1 \in L^2(0,T;\mathcal{V}_\mathcal{B})$ of the sequence $U_{1,\tau}$ solves equation* (5a) *in $\mathcal{V}_\mathcal{B}^*$, i.e., for test functions in $\mathcal{V}_\mathcal{B}$. Furthermore, U_1 has a generalized derivative, which satisfies $\dot{U}_1 \in L^2(0,T;\mathcal{V}_\mathcal{B}^*)$.*

Proof. As in [11, Ch. 4], we consider equation (15) with test functions $v \in \mathcal{V}_\mathcal{B}$ in the form

$$\frac{\mathrm{d}}{\mathrm{d}t}(\hat{U}_{1,\tau}, v) + \langle \mathcal{K} U_{1,\tau}, v\rangle = \langle \mathcal{F}_\tau, v\rangle - (\mathcal{B}^- \dot{\mathcal{G}}_\tau, v) - \langle \mathcal{K}\mathcal{B}^- \mathcal{G}_\tau, v\rangle.$$

With $\Phi \in C_0^\infty(0,T)$, we may rewrite this in integral form, i.e.,

$$\int_0^T -\big(\hat{U}_{1,\tau}, v\big)\dot{\Phi}(t) + \big\langle \mathcal{K} U_{1,\tau}, v\big\rangle \Phi(t)\,\mathrm{d}t = \int_0^T \big\langle \mathcal{F}_\tau - \mathcal{B}^- \dot{\mathcal{G}}_\tau - \mathcal{K}\mathcal{B}^- \mathcal{G}_\tau, v\big\rangle \Phi(t)\,\mathrm{d}t.$$

We advance to the limit $\tau \to 0$ and obtain for the right-hand side

$$\int_0^T \big\langle \mathcal{F}_\tau - \mathcal{B}^- \dot{\mathcal{G}}_\tau - \mathcal{K}\mathcal{B}^- \mathcal{G}_\tau, v\big\rangle \Phi(t)\,\mathrm{d}t \longrightarrow \int_0^T \big\langle \mathcal{F} - V_2 - \mathcal{K} U_2, v\big\rangle \Phi(t)\,\mathrm{d}t.$$

Furthermore, the weak convergence of $U_{1,\tau}$ and $\hat{U}_{1,\tau}$ in $L^2(0,T;\mathcal{V}_\mathcal{B})$ implies

$$\int_0^T -\big(\hat{U}_{1,\tau}, v\big)\dot{\Phi} + \big\langle \mathcal{K} U_{1,\tau}, v\big\rangle \Phi\,\mathrm{d}t \longrightarrow \int_0^T -\big(U_1, v\big)\dot{\Phi} + \big\langle \mathcal{K} U_1, v\big\rangle \Phi\,\mathrm{d}t.$$

As a result, the obtained limit $U_1 \in L^2(0,T;\mathcal{V}_\mathcal{B})$ satisfies for all $v \in \mathcal{V}_\mathcal{B}$,

$$\frac{\mathrm{d}}{\mathrm{d}t}(U_1, v) + (U_2, v) + \langle \mathcal{K}(U_1 + U_2), v\rangle = \langle \mathcal{F}, v\rangle. \tag{28}$$

It remains to show that U_1 has a generalized derivative. From the definition of $\hat{U}_{1,\tau}$ in (21) we know that its time derivative equals Du_1^j for $t \in]t_{j-1}, t_j[$ and is bounded in $L^2(0,T;\mathcal{V}_\mathcal{B}^*)$ due to (22). Thus, there exists a subsequence, which weakly converges to a limit $V_1 \in L^2(0,T;\mathcal{V}_\mathcal{B}^*)$. For every $\Phi \in C_0^\infty(0,T)$ and $v \in \mathcal{V}_\mathcal{B}$ this limit satisfies the equality

$$\begin{aligned}\int_0^T \big\langle U_1(t), v\big\rangle\ \dot{\Phi}(t)\,\mathrm{d}t &= \lim_{\tau\to 0} \int_0^T \big\langle \hat{U}_{1,\tau}(t), v\big\rangle\ \dot{\Phi}(t)\,\mathrm{d}t \\ &= \lim_{\tau\to 0} -\int_0^T \big\langle \dot{\hat{U}}_{1,\tau}(t), v\big\rangle\ \Phi(t)\,\mathrm{d}t = -\int_0^T \big\langle V_1(t), v\big\rangle\ \Phi(t)\,\mathrm{d}t.\end{aligned}$$

This shows that $\dot{U}_1 = V_1 \in L^2(0,T;\mathcal{V}_\mathcal{B}^*)$ in the generalized sense. Finally, we have to check whether U_1 satisfies the stated initial condition. Since $\hat{U}_{1,\tau} \rightharpoonup U_1$ as well as $\frac{\mathrm{d}}{\mathrm{d}t}\hat{U}_{1,\tau} \rightharpoonup \dot{U}_1 = V_1$, for $\Phi \in C^1([0,T])$ with $\Phi(T) = 0$ and arbitrary $v \in \mathcal{V}_\mathcal{B}$, we derive by the integration by parts formula that

$$0 = \lim_{\tau\to 0} \int_0^T \big\langle \dot{\hat{U}}_{1,\tau} - \dot{U}_1, v\big\rangle \Phi\,\mathrm{d}t = -\big(a_0 - U_1(0), v\big)\Phi(0).$$

Since $\mathcal{V}_\mathcal{B}$ is dense in $\overline{\mathcal{V}_\mathcal{B}}^\mathcal{H}$ by definition, this implies $U_1(0) = a_0$. □

Since we were not able to show the uniform boundedness of Λ_τ, we cannot prove its convergence to the weak solution of the operator DAE. However, we can show that the limit $\tilde{\Lambda}$ solves the operator DAE in a weaker sense. To be more precise, we state this result in the following theorem.

Theorem 5.4. *Assume $\mathcal{F} \in L^2(0,T;\mathcal{V}^*)$, $\mathcal{G} \in H^1(0,T;\mathcal{Q}^*)$, and $u_1^0 = a_0 \in \mathcal{V}_\mathcal{B}$. Then, for any sequence of step sizes with $\tau \to 0$ the sequence $\tilde{\Lambda}_\tau$ converges weakly to $\tilde{\Lambda}$ in $L^2(0,T;\mathcal{Q})$ and $(U_1, U_2, V_2, \tilde{\Lambda})$ solves system (5) in the weak distributional sense, meaning that for all $v \in \mathcal{V}$ and $\Phi \in C_0^\infty(0,T)$ it holds that*

$$\int_0^T -\big(U_1, v\big)\dot{\Phi} + \big(V_2, v\big)\Phi + \big\langle \mathcal{K}(U_1 + U_2), v\big\rangle \Phi - \big\langle \mathcal{B}^*\tilde{\Lambda}, v\big\rangle \dot{\Phi}\, dt = \int_0^T \big\langle \mathcal{F}, v\big\rangle \Phi\, dt.$$

Remark 5.2. If we assume more regularity of the data as in Remark 5.1, i.e., $\mathcal{F} \in L^2(0,T;\mathcal{H}^*)$ and the orthogonality of the decomposition $\mathcal{V} = \mathcal{V}_\mathcal{B} \oplus \mathcal{V}^\mathrm{c}$ with respect to the inner product $(\cdot,\cdot) + \langle \mathcal{K}\cdot, \cdot\rangle$, then the (weak) limits U_1, U_2, V_2, and Λ solve the regularized operator DAE (5). In addition, we have $\dot{U}_1 \in L^2(0,T;\mathcal{H})$.

Remark 5.3. The convergence of the Euler scheme for a different kind of regularization was shown in [5]. Therein, also algebraically stable Runge-Kutta schemes were analyzed.

5.4. Perturbations. In the previous subsection, we have proven the convergence of the Euler scheme if we assume that the PDEs were solved exactly in every time step. In order to prove the convergence of the Rothe method, errors due to the spatial discretization have to be included as well. For this, we consider the time-discrete systems with additional perturbations of the right-hand sides. This may then be interpreted as the error of a spatial discretization, cf. [3, Sect. 10.4].

We consider system (5), discretized by the implicit Euler scheme, with additional perturbations. Note that we still consider the linear case, cf. Section 5.1 for the assumptions on $\mathcal{K}$ and $\mathcal{B}$. The differences of the exact and perturbed solution $(\hat{u}_1^j, \hat{u}_2^j, \hat{v}_2^j, \hat{\lambda}^j)$, namely,

$$e_1^j := \hat{u}_1^j - u_1^j \in \mathcal{V}_\mathcal{B}, \quad e_2^j := \hat{u}_2^j - u_2^j \in \mathcal{V}^\mathrm{c}, \quad e_v^j := \hat{v}_2^j - v_2^j \in \mathcal{V}^\mathrm{c}, \quad e_\lambda^j := \hat{\lambda}^j - \lambda^j \in \mathcal{Q}$$

then satisfy the equations

$$De_1^j + e_v^j + \mathcal{K}\big(e_1^j + e_2^j\big) + \mathcal{B}^* e_\lambda^j = \delta^j \quad \text{in } \mathcal{V}^*, \tag{29a}$$

$$\mathcal{B}e_2^j = \theta^j \quad \text{in } \mathcal{Q}^*, \tag{29b}$$

$$\mathcal{B}e_v^j = \xi^j \quad \text{in } \mathcal{Q}^*. \tag{29c}$$

As in the previous section, the reachable results depend on the assumed smoothness of the given data. Since we want to include estimates for the Lagrange multiplier, which underlines the positive effects of the regularization from Section 3, we consider the more regular case. For this, we consider perturbations $\delta^j \in \mathcal{H}^*$ and θ^j, $\xi^j \in \mathcal{Q}^*$. Note that for an estimate of the errors e_1^j, e_2^j, and e_v^j it would be sufficient to assume $\delta^j \in \mathcal{V}^*$. Furthermore, we assume the spaces $\mathcal{V}_\mathcal{B}$ and $\mathcal{V}^\mathrm{c}$ to be orthogonal with respect to the inner product defined by $(\cdot,\cdot) + \langle \mathcal{K}\cdot, \cdot\rangle$.

By equations (29b) and (29c) we directly obtain the estimates

$$\|e_2^j\| \le C_{\mathcal{B}^-}\|\theta^j\|_{\mathcal{Q}^*}, \qquad \|e_v^j\| \le C_{\mathcal{B}^-}\|\xi^j\|_{\mathcal{Q}^*}. \tag{30}$$

For estimates on e_1^j, we may test equation (29a) by e_1^j, similarly as in Lemma 5.1. Because of the additional smoothness $\delta \in \mathcal{H}^*$ we may also test equation (29a) by De_1^j and obtain, due to the assumed orthogonality of $\mathcal{V}_\mathcal{B}$ and $\mathcal{V}^\mathrm{c}$,

$$|De_1^j|^2 + \langle \mathcal{K}e_1^j, De_1^j\rangle = \langle \delta^j, De_1^j\rangle - (e_v^j, De_1^j) + (e_2^j, De_1^j).$$

The equality $2\langle \mathcal{K}e_1^j, De_1^j\rangle = D\langle \mathcal{K}e_1^j, e_1^j\rangle + \tau\langle \mathcal{K}De_1^j, De_1^j\rangle$ then yields together with Young's inequality,

$$|De_1^j|^2 + D\langle \mathcal{K}e_1^j, e_1^j\rangle + \tau k_1 \|De_1^j\|^2 \leq 3\|\delta^j\|_{\mathcal{H}^*}^2 + 3C_{\mathrm{emb}}^2\|e_v^j\|^2 + 3C_{\mathrm{emb}}^2\|e_2^j\|^2.$$

A summation for $j = 1, \dots, k$ and a multiplication by τ finally leads to

$$k_1\|e_1^k\|^2 + \tau\sum_{j=1}^{k}|De_1^j|^2 \leq k_2\|e_1^0\|^2 + 3\tau\sum_{j=1}^{k}\Big(\|\delta^j\|_{\mathcal{H}^*}^2 + C_{\mathrm{emb}}^2\|e_2^j\|^2 + C_{\mathrm{emb}}^2\|e_v^j\|^2\Big). \tag{31}$$

Furthermore, equation (29a) and the inf-sup condition of $\mathcal{B}$ yield

$$\beta_{\mathrm{inf}}\|e_\lambda^j\|_{\mathcal{Q}} \leq \sup_{v\in\mathcal{V}^{\mathrm{c}}}\frac{\langle \mathcal{B}^*e_\lambda^j, v\rangle}{\|v\|} \leq \|\delta^j\|_{\mathcal{V}^*} + k_2\|e_1^j\| + k_2\|e_2^j\| + C_{\mathrm{emb}}^2\|e_v^j\| + C_{\mathrm{emb}}|De_1^j|.$$

Putting the estimates (30) and (31) together, we obtain with a generic constant, which we express by $\lesssim$, that

$$\begin{aligned}\tau\beta_{\mathrm{inf}}^2\sum_{j=1}^{k}\|e_\lambda^j\|_{\mathcal{Q}}^2 &\overset{(30)}{\lesssim} \tau\sum_{j=1}^{k}\big(\|\delta^j\|_{\mathcal{V}^*}^2 + \|\theta^j\|_{\mathcal{Q}^*}^2 + \|\xi^j\|_{\mathcal{Q}^*}^2\big) + \tau\sum_{j=1}^{k}\big(\|e_1^j\|^2 + |De_1^j|^2\big)\\ &\overset{(31)}{\lesssim} \|e_1^0\|^2 + \tau\sum_{j=1}^{k}\big(\|\delta^j\|_{\mathcal{H}^*}^2 + \|\theta^j\|_{\mathcal{Q}^*}^2 + \|\xi^j\|_{\mathcal{Q}^*}^2\big).\end{aligned} \tag{32}$$

To summarize the result for the Lagrange multiplier, we assume that all the perturbations are of the same order of magnitude, i.e., $\delta^j \approx \delta \in \mathcal{H}^*$, $\theta^j \approx \theta \in \mathcal{Q}^*$, and $\xi^j \approx \xi \in \mathcal{Q}^*$. Then, the piecewise constant function $E_\lambda\colon [0,T] \to \mathcal{Q}$, defined by $E_\lambda(t) = e_\lambda^j$ for $t \in]t_{j-1}, t_j]$, satisfies

$$\beta_{\mathrm{inf}}\,\|E_\lambda\|_{L^2(0,T;\mathcal{Q})} \lesssim \|e_1^0\| + \sqrt{T}\big(\|\delta\|_{\mathcal{H}^*} + \|\theta\|_{\mathcal{Q}^*} + \|\xi\|_{\mathcal{Q}^*}\big).$$

This estimate raises hope that numerical simulations can be performed in a reasonable fashion. Note that this is only true for the regularized operator DAE. For an analogous estimate for the original formulation, equation (29c) has to be replaced by $\mathcal{B}De_2^j = D\theta^j$. Thus, the perturbation ξ^j has to be replaced by the discrete derivative of θ^j, which then leads to an additional term that scales with τ^{-1} in the error estimates.

The preceding analysis reveals that, in view of numerical approximation, a Rothe discretization of the index-1 formulation, though equivalent in theory, is preferable over a Rothe discretization of the original system. We will exemplify this theoretical result in numerical tests in the following section.

6. Examples

This section is devoted to illustrate the benefits of the regularization investigated in Section 3 by means of numerical approximations of the Navier-Stokes equations. First, we revisit the numerical results presented in [4] in the context of a Rothe discretization. Second, we give a concrete example of the perturbation results from Section 5.4 by illustrating how a perturbation induced by a mesh adaption gets numerically differentiated in the index-2 but not in the index-1 formulation.

6.1. Navier-Stokes equations. We consider the standard formulation of the Navier-Stokes equations [33] for an incompressible flow in a domain $\Omega \subset \mathbb{R}^d$,

$$\dot{u} + (u \cdot \nabla)u - \nu\Delta u + \nabla p = f, \qquad \nabla \cdot u = 0. \tag{33}$$

We interpret the pressure p as a Lagrange multiplier that couples the incompressibility constraint $\nabla \cdot u = 0$ to the state equations. Then, equation (33) takes the form of system (3) with the spaces chosen as $\mathcal{V} = [H_0^1(\Omega)]^d$, $\mathcal{H} = [L^2(\Omega)]^d$, and $\mathcal{Q} = L^2(\Omega)/\mathbb{R}$.

The operator $\mathcal{B}\colon \mathcal{V} \to \mathcal{Q}^*$ is defined as the divergence and $\mathcal{K}\colon \mathcal{V} \to \mathcal{V}^*$ is the operator representing convection and diffusion, cf. Example 2.2. Generally, the nonlinearity $\mathcal{K}$ only extends to $\mathcal{K}\colon L^2(0,T;\mathcal{V}) \to L^1(0,T;\mathcal{V}^*)$, cf. [33, Lem. III.3.1]. This causes the main difficulties in the existence theory for the Navier-Stokes equations. However, this lower regularity does not affect the splitting as proposed in Section 3. In particular, Assumption 3.1 is fulfilled. The weak divergence operator is linear and bounded and there exists a continuous right inverse as shown, e.g., in [32, Lem. I.4.1]. The splitting $\mathcal{V} = \mathcal{V}_\mathcal{B} \oplus \mathcal{V}^c$ is then given by the space of divergence-free functions and its (orthogonal) complement. Note that our approach is different from [13] where the splitting of $\mathcal{V}$ is used to eliminate the constraints rather then to augment the system.

For a fixed spatial discretization and for similar approaches to the nonlinear parts, the discrete equations obtained via Rothe's method coincide with the equations stemming from the method of lines. Thus, the numerical study conducted in [4] also serves as an example for the advantages of the index-1 formulation as the base for Rothe's method.

6.2. Adaptive changes of the mesh. The benefits of the index-1 formulation become particularly apparent for space discretizations that change with time, when, e.g., the mesh is adapted to the current state of the system. Note that the opportunity to adapt the mesh between time steps is the major advantage of Rothe's method over the method of lines.

Let the superscripts $+$, c, and $-$ denote the next, current, and previous value of the variables, respectively. We use the same superscripts for the discrete operators to denote possibly different spatial discretizations. We consider the algebraic systems that are obtained from (33) after the time-discretization and, subsequently, the discretization of the space on the currently considered mesh. With the same notation as used in [4, Ch. 3.3] for the method of lines, for the original system (3), the update to (q^+, p^+) from the current iterate (q^c, p^c) via a time step of length τ is obtained via

$$\begin{bmatrix} \frac{1}{\tau}M^+ & -{B^+}^T \\ B^+ & 0 \end{bmatrix} \begin{bmatrix} q^+ \\ p^+ \end{bmatrix} = \begin{bmatrix} \frac{1}{\tau}M^+ q^c + f^+ - K^+(q^+) \\ g^+ \end{bmatrix}. \tag{34}$$

To advance by one time-step and the regularized index-1-type formulation (5), we propose the solution of

$$\begin{bmatrix} \frac{1}{\tau}M_{11}^+ & M_{12}^+ & -{B_1^+}^T & 0 \\ \frac{1}{\tau}M_{21}^+ & M_{22}^+ & -{B_2^+}^T & 0 \\ \frac{1}{\tau}B_1^+ & B_2^+ & 0 & 0 \\ B_1^+ & 0 & 0 & B_2^+ \end{bmatrix} \begin{bmatrix} q_1^+ \\ \tilde{q}_2^+ \\ p^+ \\ q_2^+ \end{bmatrix} = \begin{bmatrix} \frac{1}{\tau}M_{11}^+ q_1^c + f_1^+ - K_1^+(q_1^+, q_2^+) \\ \frac{1}{\tau}M_{21}^+ q_1^c + f_2^+ - K_2^+(q_1^+, q_2^+) \\ \frac{1}{\tau}B_1^+ q_1^c + \dot{g}^+ \\ g^+ \end{bmatrix}. \tag{35}$$

The different stability properties become evident, if one examines the inherent equation for the pressure update p^+, derived via premultiplying the upper part of the equations by $B^+(M^+)^{-1}$. In the index-2 case (34), this leads to

$$-B^+(M^+)^{-1}B^{+T}p^+ = \frac{B^+q^c - B^+q^+}{\tau} + B^+(M^+)^{-1}\big[f^+ - K(q^+)\big]. \tag{36}$$

The index-1 formulation yields for the pressure

$$-B^+(M^+)^{-1}B^{+T}p^+ = \frac{1}{\tau}B^+\begin{bmatrix} q_1^c - q_1^+ \\ -\tau\tilde{q}_2^+ \end{bmatrix} + B^+(M^+)^{-1}\big[f^+ - K(q_1^+, q_2^+)\big]. \tag{37}$$

Comparing equations (36) and (37) to the time-continuous formula for the pressure

$$-B^+(M^+)^{-1}B^{+T}p^+ = -\dot{g}^+ + B^+(M^+)^{-1}\big[f^+ - K(q^+)\big],$$

we find that the consistency errors are given as

$$e^+_{\mathrm{ind2}} := -\frac{1}{\tau}(B^+q^c - g^+) - \dot{g}^+ \quad \text{and} \quad e^+_{\mathrm{ind1}} := -\frac{1}{\tau}B^+\begin{bmatrix} q_1^c - q_1^+ \\ -\tau\tilde{q}_2^+ \end{bmatrix} - \dot{g}^+$$

for the index-2 and index-1 scheme, respectively. Unless the changes in the discretization and in B are smooth, for different meshes, i.e., $B^+q^c \neq B^cq^c = g^c$, the consistency error e^+_{ind2} will not approach zero as $\tau \to 0$. More likely, for switches in the mesh, which may appear in every time step, this term leads to an error in p^+ that scales with τ^{-1}, cf. Section 5.4. In the index-1 formulation the error term e^+_{ind1} is not present at all, since the equation $\frac{1}{\tau}B^+\begin{bmatrix} q_1^c - q_1^+ \\ -\tau\tilde{q}_2^+ \end{bmatrix} - \dot{g}^+ = 0$ is an explicit part of the numerical scheme.

Remark 6.1. The error e^+ is due to the changing meshes. In the case of inexact system solves it will add to the errors that were investigated in [4].

For a numerical example, we consider the following numerical approach to the cylinder wake at *Reynolds number* $Re = 60$ as described in [4]. The spatial component is discretized by means of *Crouzeix-Raviart* elements on a *coarse grid* with 7404 velocity and 2413 pressure nodes and on a *fine grid* with 15110 velocity and 4963 pressure nodes. The time evolution on the interval $[0, 2]$ is discretized using the *Euler* scheme – implicit in the linear parts and explicit in the nonlinearity – on a grid of 2048 and 4096 equidistant points and starting with the steady state *Stokes solution*. As reference solution we use the trajectory on the finer spatial mesh and a time discretization by the implicit trapezoidal rule.

To illustrate the error in the pressure induced by changes in the mesh and its scaling with the inverse of the time step length, we start the simulation on the *fine grid* and switch to the *coarse grid* at $t = 0.67$. At $t = 1.33$ we switch back to the fine grid. The code for this numerical example is available from the author's public *git repository* [18].

The error e^+_{ind2} and its scaling are well visible in the index-2 formulation while the pressure error at the mesh switches in the index-1 formulation is less prominent and obviously independent of the time-step size, cf. Figure 6.1(a). Therein we have plotted the pointwise in time approximation errors $\|p_{\mathrm{ref}}(t) - p_{\mathsf{Nts}}\|_{L^2(\Omega)}$, where p_{ref} is the reference solution, p_{Nts} is the numerical approximation, and Ω is the computational domain. Note, that in the index-2 case, the error e^+_{ind2} affects the pressure update only instantaneously. This is due to the implicit decoupling of pressure and velocity that makes the velocity approximation independent of the

pressure so that an error in the pressure will not spread in time. Accordingly, the large amplitude in the pressure error at the switching points is not seen in the velocity approximations, cf. Figure 6.1(b), where the pointwise temporal error in the velocity, that is defined in the same way as the pressure error, is plotted. For inexact solves, however, the implicit splitting of q and p in (34) is not exact such that a single occurrence of e^+ will spread to the velocity computation and, thus, linger in the velocity approximation forever, cf. the numerical results in [4].

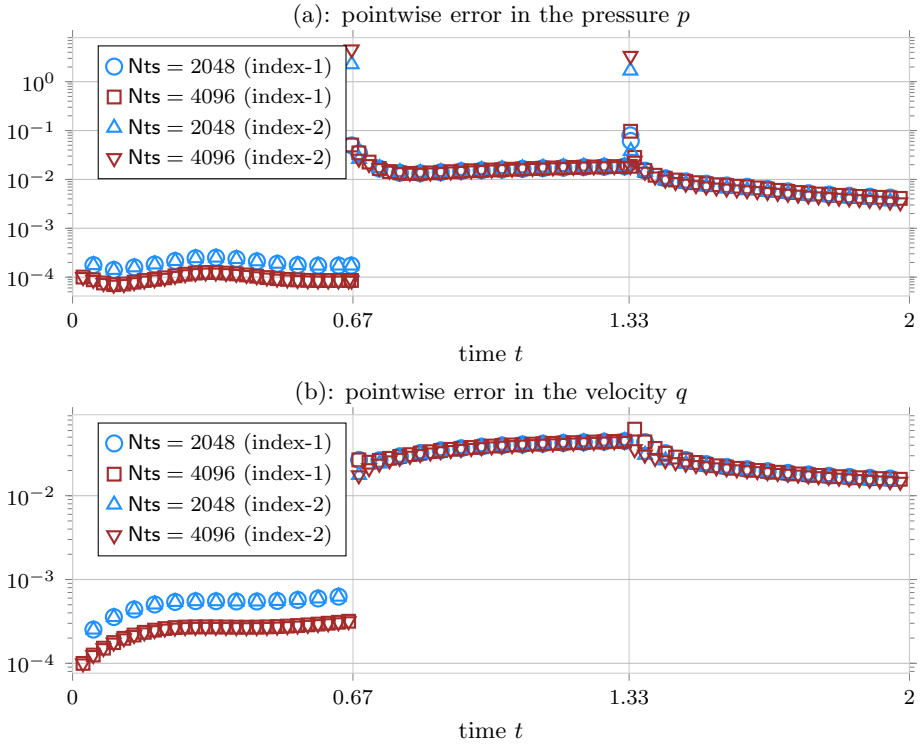

FIGURE 6.1. Pointwise in time error for q and p for the index-1 and index-2 formulations for various number of time steps Nts and with mesh switches at $t = 0.67$ and $t = 1.33$ plotted at every 50-th point of the temporal grid and at the points where the switches occur.

7. Conclusion

Within this paper, we have introduced a reformulation for a special class of semi-explicit operator DAEs with linear constraints such that a standard spatial discretization by finite elements leads to a DAE of index 1, rather than index 2. Thus, the procedure can be seen as an index reduction or regularization for operator DAEs.

Furthermore, we have proven the convergence of the Rothe discretization on the base of the implicit Euler scheme. We have quantified the advantages of the reformulation over the original schemes in terms of stability estimates concerning

the robustness against perturbations in the right-hand sides. Particularly, we have shown that derivatives of perturbations, that may occur in the original formulation, are not present in the solutions of the reformulated equations.

Finally, we have illustrated the advantages of the regularized formulation in a numerical simulation of flow equations where the spatial discretization changes at certain time points.

Acknowledgments

The work of R. Altmann was supported by the ERC Advanced Grant MODSIMCONMP and the Berlin Mathematical School BMS.

References

[1] R. A. Adams and J. J. F. Fournier, *Sobolev Spaces*, Elsevier, Amsterdam, 2003.
[2] R. Altmann, *Index reduction for operator differential-algebraic equations in elastodynamics*, Z. Angew. Math. Mech. (ZAMM), 93 (2013), pp. 648–664.
[3] ———, *Regularization and Simulation of Constrained Partial Differential Equations*, PhD thesis, Technische Universität Berlin, 2015.
[4] R. Altmann and J. Heiland, *Finite element decomposition and minimal extension for flow equations*, ESAIM Math. Model. Numer. Anal., 49 (2015), pp. 1489–1509.
[5] R. Altmann and C. Zimmer, *Runge-Kutta methods for linear semi-explicit operator differential-algebraic equations*, Math. Comp., 87 (2018), p. 149–174.
[6] M. Arnold and B. Simeon, *Pantograph and catenary dynamics: A benchmark problem and its numerical solution*, Appl. Numer. Math., 34 (2000), pp. 345–362.
[7] D. Braess, *Finite Elements - Theory, Fast Solvers, and Applications in Solid Mechanics*, Cambridge University Press, New York, third ed., 2007.
[8] K. Brenan, S. Campbell, and L. R. Petzold, *Numerical solution of initial-value problems in differential-algebraic equations*, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1996.
[9] S. L. Campbell and W. Marszalek, *The index of an infinite-dimensional implicit system*, Math. Comput. Model. Dyn. Syst., 5 (1999), pp. 18–42.
[10] 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 (1973), pp. 33–75.
[11] E. Emmrich, *Analysis von Zeitdiskretisierungen des inkompressiblen Navier–Stokes-Problems*, Cuvillier Verlag, Göttingen, Germany, 2001.
[12] ———, *Gewöhnliche und Operator-Differentialgleichungen*, Vieweg, Wiesbaden, Germany, 2004.
[13] E. Emmrich and V. Mehrmann, *Operator differential-algebraic equations arising in fluid dynamics*, Comp. Methods Appl. Math., 13 (2013), pp. 443–470.
[14] L. C. Evans, *Partial Differential Equations*, American Mathematical Society (AMS), Providence, second ed., 1998.
[15] V. Girault and P.-A. Raviart, *Finite Element Methods for Navier–Stokes Equations. Theory and Algorithms*, Springer, Berlin, Germany, 1986.

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

[17] J. Heiland, *Decoupling and optimization of differential-algebraic equations with application in flow control*, PhD thesis, Technische Universität Berlin, 2014.

[18] ———, *TayHoodMinExtForFlowEqns*. Public Git Repository, 2016. Solution of time-dependent 2D nonviscous flow with nonconforming minimal extension, `https://github.com/highlando/TayHoodMinExtForFlowEqns`.

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

[20] ———, *Differential-Algebraic Equations. Analysis and Numerical Solution*, European Mathematical Society Publishing House, Zürich, Switzerland, 2006.

[21] R. Lamour, R. März, and C. Tischendorf, *Differential-Algebraic Equations: A Projector Based Analysis*, Differential-Algebraic Equations Forum, Springer, Germany, 2013.

[22] W. S. Martinson and P. I. Barton, *A differentiation index for partial differential-algebraic equations*, SIAM J. Sci. Comput., 21 (2000), pp. 2295–2315.

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

[24] V. Mehrmann, *Index concepts for differential-algebraic equations*, in Encyclopedia of Applied and Computational Mathematics, T. Chan, W. Cook, E. Hairer, J. Hastad, A. Iserles, H. Langtangen, C. Le Bris, P. Lions, C. Lubich, A. Majda, J. McLaughlin, R. Nieminen, J. Oden, P. Souganidis, and A. Tveito, eds., Springer-Verlag, Berlin, 2013.

[25] J. Rang and L. Angermann, *Perturbation index of linear partial differential-algebraic equations*, Appl. Numer. Math., 53 (2005), pp. 437–456.

[26] M. Renardy and R. C. Rogers, *An Introduction to Partial Differential Equations*, Springer-Verlag, New York, second ed., 2004.

[27] E. Rothe, *Zweidimensionale parabolische Randwertaufgaben als Grenzfall eindimensionaler Randwertaufgaben*, Math. Ann., 102 (1930), pp. 650–670.

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

[29] M. Ružička, *Nichtlineare Funktionalanalysis: Eine Einführung*, Springer, UK, 2004.

[30] B. Simeon, *Modelling a flexible slider crank mechanism by a mixed system of DAEs and PDEs*, Math. Comp. Model. Dyn., 2 (1996), pp. 1–18.

[31] ———, *On Lagrange multipliers in flexible multibody dynamics*, Comput. Methods Appl. Mech. Engrg., 195 (2006), pp. 6993–7005.

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

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

[34] C. Tischendorf, *Coupled systems of differential algebraic and partial differential equations in circuit and device simulation. Modeling and numerical analysis*, habilitationsschrift, Technische Universität Berlin, 2004.

[35] 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.

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

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

E-mail: robert.altmann@math.uni-augsburg.de

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

E-mail: heiland@mpi-magdeburg.mpg.de