AN ULTRAWEAK VARIATIONAL METHOD FOR PARAMETERIZED LINEAR DIFFERENTIAL-ALGEBRAIC EQUATIONS

EMIL BEURER, MORITZ FEUERLE, NIKLAS REICH, AND KARSTEN URBAN

ABSTRACT. We investigate an ultraweak variational formulation for (parameterized) linear differential-algebraic equations (DAEs) w.r.t. the time variable which yields an optimally stable system. This is used within a Petrov-Galerkin method to derive a certified detailed discretization which provides an approximate solution in an ultraweak setting as well as for model reduction w.r.t. time in the spirit of the Reduced Basis Method (RBM). A computable sharp error bound is derived. Numerical experiments are presented that show that this method yields a significant reduction and can be combined with well-known system theoretic methods such as Balanced Truncation to reduce the size of the DAE.

1. INTRODUCTION

Differential-Algebraic Equations (DAEs) are widely used to model several processes in science, engineering, medicine and other fields. Theory and numerical approximation methods have intensively been studied in the literature, see e.g. [5, 14, 18, 29], or [13, 24], which are the first two books in a forum series on DAEs. Quite often, the dimension of DAEs modeling realistic problems is so large that an efficient numerical solution (in particular in realtime environments or within optimal control) is impossible. To address this issue, Model Order Reduction (MOR) techniques have been developed and successfully applied. There is a huge amount of literature, we just mention [3, 4, 8, 20, 25, 26].

All methods described in those references address a reduction of the dimension of the system, whereas the temporal discretization is untouched. This paper starts at this point. We have been working on space-time variational formulations for (parameterized) partial differential equations (PPDEs) over the last decade. One particular issue has been the stability of the arising discretization which admits tight error-residual relations and thus builds the backbone for model reduction. It turns out that an *ultraweak* formulation is the right tool to achieve this goal. In [10], we have used this framework for deriving an optimally stable variational formulation of linear time-invariant systems (LTIs). In this paper, we extend the ultraweak framework to (parameterized) DAEs and show that this can be combined with system theoretic methods such as Balanced Truncation (BT, [25]) to derive a reduction in the system dimension and time discretization size.

1.1. **Differential-algebraic equations (DAEs).** Let $E, A \in \mathbb{R}^{n\times n}$, $n \in \mathbb{N}$, be two matrices (E is typically singular), $I = (0, T)$, $T > 0$, a time interval, $x_0 \in \mathbb{R}^n$ some initial value and $f : I \to \mathbb{R}^n$ a given right-hand side. Then, we are interested in

1991 *Mathematics Subject Classification.* Primary 34A09, 65L80, 65M60.

The authors have no competing interests to declare that are relevant to the content of this article. We acknowledge support by the state of Baden-Württemberg through bwHPC.

the solution $x : I \to \mathbb{R}^n$ (the state) of the following initial value problem of a linear differential-algebraic equation (DAE) with constant coefficients

$$E\dot{x}(t) - Ax(t) = f(t), \quad \forall t \in I, \qquad x(0) = x_0.$$

In order to ensure well-posedness (in an appropriate manner), we shall always assume that the initial value x_0 is *consistent* with the right-hand side f, which means that there exists some $\hat{x}_0 \in \mathbb{R}^n$ such that $E\hat{x}_0 - Ax_0 = \lim_{t\to 0+} f(t)$ holds. Finally, we assume that the matrix pencil $\{E, -A\}$ is regular (i.e. $\det(\lambda E - A) \neq 0$ for some $\lambda \in \mathbb{R}$) with index $\operatorname{ind}\{E, -A\} =: k \in \mathbb{N}$, [12].[1]

1.2. **Parameterized DAEs (PDAEs).** We are particularly interested in the situation, where one does not only have to solve the above DAE once, but several times and highly efficient (e.g. in realtime, optimal control or cold computing devices) for different data. In order to describe that situation, we are considering a *parameterized* DAE (PDAE) as follows. For some parameter vector $\mu \in \mathcal{P}$, $\mathcal{P} \subset \mathbb{R}^P$ being a compact set, we are seeking $x_\mu : I \to \mathbb{R}^n$ such that

$$E\,\dot{x}_\mu(t) - A_\mu\, x_\mu(t) = f_\mu(t), \quad \forall t \in I, \qquad x_\mu(0) = x_{0,\mu}, \tag{1.1}$$

where A_μ, f_μ and $x_{0,\mu}$ are a parameter-dependent matrix, a right-hand side and an initial condition, respectively, whereas E is assumed to be independent of μ, see below. In order to be able to solve such a PDAE highly efficient for many parameters, it is quite standard to assume that parameters and variables can be separated, see e.g. [17]. This is done by assuming a so-called *affine decomposition* of the data, i.e., E is (for simplicity of exposition) assumed to be parameter-independent and

$$A_\mu = \sum_{q=1}^{Q_A} \vartheta_q^A(\mu)\, \tilde{A}_q, \qquad f_\mu(t) = \sum_{q=1}^{Q_f} \vartheta_q^f(\mu)\, \tilde{f}_q(t), \qquad x_{0,\mu} = \sum_{q=1}^{Q_x} \vartheta_q^x(\mu)\, \tilde{x}_{0,q}. \tag{1.2}$$

If such a decomposition is not given, we may produce an affinely decomposed approximation by means of the *(Discrete) Empirical Interpolation Method* (*(D)EIM*, [2, 7]; see also [8] for a system theoretic MOR for such PDAEs). For well-posedness, we assume that the matrix pencil $\{E, -A_\mu\}$ is regular with index $\operatorname{ind}\{E, -A_\mu\} = k_\mu$ for all $\mu \in \mathcal{P}$.

1.3. **Reduction to homogeneous initial conditions.** Using some standard arguments, (1.1) can be reduced to homogeneous initial conditions $x_\mu(0) = 0$. To this end, construct some smooth extension of the initial data $\bar{x}_\mu \in C^1(\bar{I})^n$, $\bar{x}_\mu(0) = x_{0,\mu}$. Then, let $\hat{x}_\mu : I \to \mathbb{R}^n$ solve (1.1) with f_μ replaced by $\hat{f}_\mu := f_\mu - E\dot{\bar{x}}_\mu + A_\mu\bar{x}_\mu$ and homogeneous initial condition $\hat{x}_\mu(0) = 0$. Then, $x_\mu := \hat{x}_\mu + \bar{x}_\mu$ solves the original problem (1.1). If the PDAE possess an affine decomposition (1.2), it is readily seen that the modified right-hand side $\hat{f}_\mu$ also admits an affine decomposition. Hence, we can always restrict ourselves to the case of homogeneous initial conditions $x_\mu(0) = 0$, keeping in mind that variable initial conditions can be realized by different right-hand sides.

[1] Each regular matrix pencil can be transformed into *Weierstrass-Kronecker canonical form* $P(\lambda E - A)Q = \operatorname{diag}(\lambda Id - W, \lambda N - Id)$ with regular matrices $P, Q \in \mathbb{C}^{n\times n}$, [11]. The index of a regular matrix pencil $\{E, -A\}$ is then defined by $\operatorname{ind}\{E, -A\} := \operatorname{ind}\{N\} := \min\{k \in \mathbb{N} : N^k = 0\}$.

1.4. **Organization of the material.** The remainder of this paper is organized as follows. In Section 2, we derive an ultraweak variational formulation of (1.1) and prove its well-posedness. Section 3 is devoted to a corresponding Petrov-Galerkin discretization and the numerical solution, which is then used in Section 4 to derive a certified reduced model. In Section 5, we report results of our numerical experiments and end with conclusions and an outlook in Section 6.

2. An ultraweak variational formulation

It is well-known that, for any fixed parameter $\mu \in \mathcal{P}$, the problem (1.1) admits a unique classical solution $x_\mu \in C^{k_\mu}(\bar{I})^n$ for consistent initial conditions provided that $f_\mu \in C^{k_\mu-1}(\bar{I})^n$, e.g. [18, Lemma 2.8.]. This is a severe regularity assumption, which is one of the reasons why we are interested in a variational formulation. As we shall see, an *ultraweak* setting is appropriate in order to prove well-posedness, in particular stability. It turns out that this setting is also particularly useful for model reduction of (1.1) w.r.t. the time variable in the spirit of the reduced basis method, see §4 below.

2.1. **Ultraweak formulation of PPDEs.** In order to describe an ultraweak variational formulation for the above PDAE, we will review such formulations for parametric *partial* differential equations (PPDEs). In particular, we are going to follow [9] in which well-posed (ultraweak) variational forms for transport problems have been introduced, see also [6, 16, 30]. We will then transfer this framework to PDAEs in §2.2.

Let $\Omega \subset \mathbb{R}^n$ be some open and bounded domain. We consider a *classical*[2] linear operator $B_{\mu;\circ}$ on Ω with classical domain

$$D(B_{\mu;\circ}) = \{x \in C(\bar{\Omega}) : x_{|\partial\Omega} = 0, B_{\mu;\circ}x \in C(\Omega)\}$$

and aim at solving

$$B_{\mu;\circ}x_\mu = f_\mu \text{ (pointwise) on } \Omega, \qquad x_{\mu_{|\partial\Omega}} = 0. \tag{2.1}$$

Note that the definition of $B_{\mu;\circ}$ also incorporates essential homogeneous boundary conditions (in case of a PDAE described below this is the initial condition, which is independent of the parameter). Let $\{B^*_{\mu;\circ}, D(B^*_{\mu;\circ})\}$ denote the operator, which is adjoint to $\{B_{\mu;\circ}, D(B_{\mu;\circ})\}$, i.e., $B^*_{\mu;\circ}$ is defined as the formal adjoint of $B_{\mu;\circ}$ by $(B_{\mu;\circ}x, y)_{L_2(\Omega)} = (x, B^*_{\mu;\circ}y)_{L_2(\Omega)}$ for all $x, y \in C^\infty_0(\Omega)$ and its domain $D(B^*_{\mu;\circ})$ which includes the corresponding adjoint essential boundary conditions (so that the above equation still holds true for all $x \in D(B_{\mu;\circ})$, $y \in D(B^*_{\mu;\circ})$). Denoting the range of an operator B by $R(B)$, we have $B_{\mu;\circ} : D(B_{\mu;\circ}) \to R(B_{\mu;\circ})$ and $B^*_{\mu;\circ} : D(B^*_{\mu;\circ}) \to R(B^*_{\mu;\circ})$. The following assumptions[3] turned out to be crucial for ensuring the well-posedness:

(B1) $D(B_{\mu;\circ}), D(B^*_{\mu;\circ}), R(B^*_{\mu;\circ}) \subseteq L_2(\Omega)$ with all embeddings being dense;
(B2) $B^*_{\mu;\circ}$ is injective on $D(B^*_{\mu;\circ})$.

Due to (B2), the injectivity of the adjoint operator, the following quantity

$$|||\cdot|||_\mu := \|B^*_\mu \cdot\|_{L_2(\Omega)}$$

is a norm on $D(B^*_{\mu;\circ})$, where B^*_μ is to be understood as the continuous extension of $B^*_{\mu;\circ}$ onto Y_μ, i.e., $B^*_\mu : Y_\mu \to L_2(\Omega)$, where

$$Y_\mu := \operatorname{clos}_{|||\cdot|||_\mu}(D(B^*_{\mu;\circ})), \quad (v, w)_{Y_\mu} := (B^*_\mu v, B^*_\mu w)_{L_2(\Omega)}, \quad \|v\|^2_{Y_\mu} := (v, v)_{Y_\mu} = |||v|||^2_\mu,$$

[2] By *classical* we mean defined in a pointwise manner.

[3] The framework in [9] is slightly more general.

is a Hilbert space. Defining the bilinear form

$$b_\mu : L_2(\Omega) \times Y_\mu \to \mathbb{R} \quad \text{by} \quad b_\mu(x,y) := (x, B_\mu^* y)_{L_2(\Omega)},$$

yields an ultraweak form of (2.1): For $f \in Y_\mu'$[4], determine $x_\mu \in L_2(\Omega)$ such that

$$b_\mu(x_\mu, y_\mu) = f_\mu(y_\mu) \qquad \forall y_\mu \in Y_\mu. \tag{2.2}$$

Well-posedness including optimal stability is now ensured:

Lemma 2.1. *Problem* (2.2) *has a unique solution* $x_\mu \in L_2(\Omega)$ *and is optimally stable, i.e.,* $\gamma_\mu = \beta_\mu = \beta_\mu^* = 1$*, where the continuity constant is defined as*

$$\gamma_\mu := \sup_{x \in L_2(\Omega)} \sup_{y_\mu \in Y_\mu} \frac{b_\mu(x, y_\mu)}{\|x\|_{L_2(\Omega)} \|y_\mu\|_{Y_\mu}},$$

and primal resp. dual inf-sup constants read

$$\beta_\mu := \inf_{x \in L_2(\Omega)} \sup_{y_\mu \in Y_\mu} \frac{b_\mu(x, y_\mu)}{\|x\|_{L_2(\Omega)} \|y_\mu\|_{Y_\mu}}, \qquad \beta_\mu^* := \inf_{y_\mu \in Y_\mu} \sup_{x \in L_2(\Omega)} \frac{b_\mu(x, y_\mu)}{\|x\|_{L_2(\Omega)} \|y_\mu\|_{Y_\mu}}.$$

Proof. See [9, Proposition 3.1 and Corollary 3.2]. □

2.2. **An ultraweak formulation of PDAEs.** We are now going to apply the framework of §2.1 to the classical form (1.1) of the PDAE. Again, w.l.o.g. we restrict ourselves to homogeneous initial conditions $x_\mu(0) = 0$, as stated in §1.3.

It is immediate that we can generalize ultraweak formulations for scalar-valued functions in $L_2(\Omega)$ as above to systems, i.e., $L_2(\Omega)^n \equiv L_2(\Omega; \mathbb{R}^n)$. For PDAEs, we choose $L_2(I)^n$ with the inner product $(\cdot,\cdot)_{L_2} \equiv (\cdot,\cdot)_{L_2(I)^n}$, whereas $(\cdot,\cdot)$ denotes the Euclidean inner product of vectors. The linear operator $\{B_{\mu;\circ}, D(B_{\mu;\circ})\}$ corresponding to (1.1) reads

$$B_{\mu;\circ} := E\frac{d}{dt} - A_\mu, \qquad D(B_{\mu;\circ}) := \{x \in C^{k_\mu}(I)^n \cap C(\bar{I})^n \,:\, x(0) = 0\}.$$

The formal adjoint operator $B_{\mu;\circ}^*$ is easily derived by integration by parts, i.e.,

$$\begin{aligned}(B_{\mu;\circ}x, y)_{L_2} &= (E\dot{x} - A_\mu x, y)_{L_2} = (\dot{x}, E^T y)_{L_2} - (x, A_\mu^T y)_{L_2} \\ &= (x(T), E^T y(T)) - (x(0), E^T y(0)) - (x, E^T \dot{y})_{L_2} - (x, A_\mu^T y)_{L_2} \\ &= (x, -E^T \dot{y} - A_\mu^T y)_{L_2} =: (x, B_{\mu;\circ}^* y)_{L_2} \quad \forall x, y \in C_0^\infty(I)^n,\end{aligned}$$

which shows that

$$B_{\mu;\circ}^* := -E^T \frac{d}{dt} - A_\mu^T,$$

$$D(B_{\mu;\circ}^*) \equiv C_E^1(I)^n := \{y \in C^1(I)^n \cap C(\bar{I})^n \,:\, y(T) \in \ker(E^T)\}. \tag{2.3}$$

In fact, $(B_{\mu;\circ}x, y)_{L_2} = (x, B_{\mu;\circ}^* y)_{L_2}$ for all $x \in D(B_{\mu;\circ})$ and $y \in D(B_{\mu;\circ}^*)$ since the boundary terms above still vanish thanks to $x(0) = 0$ and $y(T) \in \ker(E^T)$. Moreover

$$R(B_{\mu;\circ}) = C^{k_\mu - 1}(I)^n \cap C(\bar{I})^n, \qquad R(B_{\mu;\circ}^*) = C(\bar{I})^n.$$

Lemma 2.2. *We have* $D(B_{\mu;\circ}), D(B_{\mu;\circ}^*), R(B_{\mu;\circ}^*) \subset L_2(I)^n$ *with dense embeddings.*

[4] Y_μ' denotes the dual space of Y_μ w.r.t. the pivot space $L_2(\Omega)$.

Proof. By the definition of $H_0^1(I)^n$ and [1, Cor. 7.24] (for $H^1(I)^n$ instead of $H^1(I)$ there, which is a trivial extension), we have

$$C_0^\infty(I)^n \subset H_0^1(I)^n \subset C(\bar{I})^n, \quad \text{hence} \quad C_0^\infty(I)^n = C_0^\infty(I)^n \cap C(\bar{I})^n.$$

With that, $C_0^\infty(I)^n \subseteq D(B_{\mu;\circ}), D(B_{\mu;\circ}^*), R(B_\circ^*) \subset L_2(I)^n$ is easy to see. Since $C_0^\infty(I)^n$ is dense in $L_2(I)^n$, its supersets $D(B_{\mu;\circ}), D(B_{\mu;\circ}^*), R(B_{\mu;\circ}^*)$ are also dense in $L_2(I)^n$. □

The above lemma ensures assumption (B1). Next, we consider (B2).

Lemma 2.3. *The adjoint operator $\{B_{\mu;\circ}^*, D(B_{\mu;\circ}^*)\}$ is injective, i.e., for $y_\mu, z_\mu \in D(B_{\mu;\circ}^*)$ with $B_{\mu;\circ}^* y_\mu = B_{\mu;\circ}^* z_\mu$ we have $y_\mu = z_\mu$.*

Proof. Setting $d_\mu := y_\mu - z_\mu$, we get $B_{\mu;\circ}^* d_\mu = 0$ and

$$-E^T \dot{d}_\mu(t) - A_\mu^T d_\mu(t) = 0, \quad \forall t \in I, \qquad d_\mu(T) = y_\mu(T) - z_\mu(T) \in \ker(E^T).$$

Due to regularity of $\{E, -A_\mu\}$ (and thus also of $\{-E^T, -A_\mu^T\}$), there are regular matrices $P_\mu, Q_\mu \in \mathbb{C}^{n\times n}$, which allow us to transform the problem into Weierstrass-Kronecker normal form, [14, 18], i.e.,

$$P_\mu E^T Q_\mu = \begin{pmatrix} Id_m & 0 \\ 0 & N_\mu \end{pmatrix}, \quad P_\mu A_\mu^T Q_\mu = \begin{pmatrix} R_\mu & 0 \\ 0 & Id_{n-m} \end{pmatrix}, \quad Q_\mu^{-1} d_\mu(t) = \begin{pmatrix} u_\mu(t) \\ v_\mu(t) \end{pmatrix},$$

where $Id_n \in \mathbb{R}^{n\times n}$ is the identity and N_μ is a nilpotent matrix with nilpotency index k_μ. This yields the equivalent representation

$$\dot{u}_\mu(t) + R_\mu u_\mu(t) = 0, \qquad \forall t \in I, \tag{2.4a}$$

$$N_\mu \dot{v}_\mu(t) + v_\mu(t) = 0, \qquad \forall t \in I, \tag{2.4b}$$

$$Q_\mu \begin{pmatrix} u_\mu(T) \\ v_\mu(T) \end{pmatrix} \in \ker(E^T). \tag{2.4c}$$

The ODE (2.4a) has the general solution $u_\mu(t) = u_\mu(T)\, e^{-R_\mu(T-t)}$. By (2.4c) we get

$$E^T Q_\mu \begin{pmatrix} u_\mu(T) \\ v_\mu(T) \end{pmatrix} = 0 = P_\mu E^T Q_\mu \begin{pmatrix} u_\mu(T) \\ v_\mu(T) \end{pmatrix} = \begin{pmatrix} Id_m & 0 \\ 0 & N_\mu \end{pmatrix} \begin{pmatrix} u_\mu(T) \\ v_\mu(T) \end{pmatrix} = \begin{pmatrix} u_\mu(T) \\ N_\mu v_\mu(T) \end{pmatrix},$$

so that $u_\mu(T) = 0$ and hence $u_\mu(t) = u_\mu(T)\, e^{-R_\mu(T-t)} = 0$ for all $t \in I$.

The initial value problem $N_\mu \dot{v}_\mu(t) + v_\mu(t) = q_\mu(t)$, $t \in I$, $v_\mu(T) = v_{\mu,T}$ with some $q_\mu \in C^{k_\mu-1}(\bar{I})^{n-m}$ has the unique solution $v_\mu(t) = \sum_{i=0}^{k_\mu-1} (-1)^i N_\mu^i q_\mu^{(i)}$, if the initial value $v_{\mu,T}$ is consistent, see e.g. [5]. We apply this for $q_\mu \equiv 0 \in C^{k_\mu-1}(\bar{I})^{n-m}$. Then, by the solution formula, we get $v_\mu \equiv 0$, since the initial value in (2.4c) is by definition trivially consistent. This yields $d_\mu \equiv 0$, i.e., $y_\mu = z_\mu$. □

Hence, we set $|||\cdot|||_\mu := \|B_\mu^* \cdot\|_{L_2}$ and choose trial and test spaces as

$$X := L_2(I)^n, \quad Y_\mu := \operatorname{clos}_{|||\cdot|||_\mu}\big(C_E^1(I)^n\big), \quad b_\mu(x,y) := (x, B_\mu^* y)_{L_2}, \tag{2.5}$$

see (2.3) and obtain the following result.

Lemma 2.4. *Under the above assumptions, we have for all $\mu \in \mathcal{P}$ that $Y_\mu \equiv Y$, where*

$$Y := H_E^1(I)^n := \{v \in H^1(I)^n : v(T) \in \ker(E^T)\}.$$

Proof. Clearly $C_E^1(I)^n \subset H_E^1(I)^n$, so that $Y_\mu \subseteq Y$ for all $\mu \in \mathcal{P}$. Now, let $y \in Y = H_E^1(I)^n$, then, by density, there is a sequence $(y_\ell)_{\ell\in\mathbb{N}} \subset C_E^1(I)^n$ such that $\|y_\ell - y\|_{H^1(I)^n} \to 0$ as $\ell \to 0$. Since $\mathcal{P}$ is compact, we have that

$$|||y_\ell - y|||_\mu = \|E^T(\dot{y}_\ell - \dot{y}) + A_\mu^T(y_\ell - y)\|_{L_2} \leq \max\{\|E\|, \|A_\mu\|\}\, \|y_\ell - y\|_{H^1(I)^n} \to 0$$

as $\ell \to \infty$. Hence, $y \in \operatorname{clos}_{|||\cdot|||_\mu}\big(C_E^1(I)^n\big) = Y_\mu$, i.e. $Y \subseteq Y_\mu$. □

The latter result must be properly interpreted. It says that Y_μ and Y coincide as sets. However, the norm $|||\cdot|||_\mu$ (and thus the topology) still depends on the parameter. The same holds true for the dual space Y' of Y induced by the L_2-inner product and normed by

$$|||f|||'_\mu := \sup_{y\in Y} \frac{(f,y)_{L_2}}{|||y|||_\mu}.$$

In particular, we have a generalized Cauchy-Schwarz inequality $(f,y)_{L_2} \leq |||f|||'_\mu\, |||y|||_\mu$.

Lemma 2.5. *Let $f_\mu \in Y'$. Then, there exists a unique weak solution $x_\mu \in X$ of*

$$b_\mu(x_\mu, y) = f_\mu(y), \qquad \forall y \in Y. \tag{2.6}$$

If (1.1) *admits a classical solution, then it coincides with x_μ. Moreover, $\gamma_\mu = \beta_\mu = \beta_\mu^* = 1$ for the constants defined in Lemma 2.1.*

Proof. The existence of a unique solution $x_\mu \in X$ (as well as $\gamma_\mu = \beta_\mu = \beta_\mu^* = 1$) is an immediate consequence of Lemma 2.1. It only remains to show that x_μ satisfying (2.6) is a weak solution of (1.1). To this end, let $\tilde{f}_\mu \in C(I)^n$ be given such that there exists a classical solution $\tilde{x}_\mu \in C^1(\bar{I})^n$ with $B_{\mu;\circ}\tilde{x}_\mu(t) = \tilde{f}_\mu(t), \forall t \in I$ and $\tilde{x}_\mu(0) = 0$. Then, define $f_\mu \in Y'$ by $f_\mu(y) := (\tilde{f}_\mu, y)_{L_2}$. We need to show that the classical solution $\tilde{x}_\mu$ of (1.1) is also the unique solution of (2.6). First, for $y \in C_E^1(I)^n$, integration by parts yields $b_\mu(\tilde{x}_\mu, y) - f_\mu(y) = (\tilde{x}_\mu, B_\mu^* y)_{L_2} - f_\mu(y) = (B_{\mu;\circ}\tilde{x}_\mu - \tilde{f}_\mu, y)_{L_2} = 0$. Second, let $y \in Y \setminus C_E^1(I)^n$, then there is $(\tilde{y}_\ell)_{\ell\in\mathbb{N}} \subset C_E^1(I)^n$ converging to y in Y, i.e., $\lim_{\ell\to\infty} |||y - \tilde{y}_\ell|||_\mu = 0$. Then, by the generalized Cauchy-Schwarz inequality

$$\begin{aligned}
|b_\mu(\tilde{x}_\mu, y) - f_\mu(y)| &= |b_\mu(\tilde{x}_\mu, y) - f_\mu(y) - b_\mu(\tilde{x}_\mu, \tilde{y}_\ell) + f_\mu(\tilde{y}_\ell)| \\
&= |(\tilde{x}_\mu, B_\mu^*(y - \tilde{y}_\ell))_{L_2} - f_\mu(y - \tilde{y}_\ell)| \\
&\leq \|\tilde{x}_\mu\|_{L_2} \|B_\mu^*(y - \tilde{y}_\ell)\|_{L_2} + |||f_\mu|||'_\mu\, |||y - \tilde{y}_\ell|||_\mu \\
&= \big(\|\tilde{x}_\mu\|_{L_2} + |||f_\mu|||'_\mu\big)\, |||y - \tilde{y}_\ell|||_\mu \to 0 \quad \text{as } \ell \to \infty,
\end{aligned}$$

so that (2.6) holds for $\tilde{x}_\mu$. □

For the ultraweak PDAE (2.6), we need a right-hand side $f_\mu \in Y'$. However, typically, the right-hand side is given within context of (1.1) as a function of time, i.e., $g_\mu : I \to \mathbb{R}^n$. Then, we simply define $f_\mu \in Y'$ by

$$f_\mu(y) := (g_\mu, y)_{L_2} = \int_I (g_\mu(t), y(t))\, \mathrm{d}t, \quad y \in Y. \tag{2.7}$$

3. Petrov-Galerkin discretization

The next step towards a numerical method for solving an ultraweak operator equation is to introduce finite-dimensional trial and test spaces yielding a Petrov-Galerkin discretization. In this section, we shall first review Petrov-Galerkin methods in general terms and then detail the specification for PDAEs.

3.1. **Petrov-Galerkin method.** In order to determine a numerical approximation, we are going to construct an appropriate finite-dimensional trial space $X_\mu^{\mathcal{N}} \subset X = L_2(I)^n$ and a parameter-independent test space $Y^{\mathcal{N}} \subset Y$ of finite (but possibly large) dimension $\mathcal{N} \in \mathbb{N}$. Then, we are seeking $x_\mu^{\mathcal{N}} \in X_\mu^{\mathcal{N}}$ such that

$$b_\mu(x_\mu^{\mathcal{N}}, y^{\mathcal{N}}) = f_\mu(y^{\mathcal{N}}), \qquad \forall y^{\mathcal{N}} \in Y^{\mathcal{N}}, \tag{3.1}$$

which leads to solving a linear system of equations $\boldsymbol{B}_\mu^{\mathcal{N}} \boldsymbol{x}_\mu^{\mathcal{N}} = \boldsymbol{f}_\mu^{\mathcal{N}}$ in $\mathbb{R}^{\mathcal{N}}$.

Remark 3.1. (a) If one would choose a discretization with $\dim(Y^{\mathcal{N}}) > \dim(X_\mu^{\mathcal{N}})$, one would need to solve a least squares problem $\|\boldsymbol{B}_\mu^{\mathcal{N}} \boldsymbol{x}_\mu^{\mathcal{N}} - \boldsymbol{f}_\mu^{\mathcal{N}}\|^2 \to \min$.

(b) If one defines the trial space according to $X_\mu^{\mathcal{N}} := B_\mu^* Y^{\mathcal{N}}$, then it is easily seen that the discrete problem (3.1) is well-posed and optimally conditioned, [6], i.e.,

$$\gamma_\mu^{\mathcal{N}} := \sup_{x \in X_\mu^{\mathcal{N}}} \sup_{y \in Y^{\mathcal{N}}} \frac{b_\mu(x,y)}{\|x\|_{L_2} |||y|||_\mu} = 1,$$

$$\beta_\mu^{\mathcal{N}} := \inf_{x \in X_\mu^{\mathcal{N}}} \sup_{y \in Y^{\mathcal{N}}} \frac{b_\mu(x,y)}{\|x\|_{L_2} |||y|||_\mu} = 1, \qquad \beta_\mu^{*,\mathcal{N}} := \inf_{y \in Y^{\mathcal{N}}} \sup_{x \in X_\mu^{\mathcal{N}}} \frac{b_\mu(x,y)}{\|x\|_{L_2} |||y|||_\mu} = 1.$$

(c) The Xu-Zikatanov lemma ([31]) ensures that the Petrov-Galerkin error is comparable with the error of the best approximation, namely

$$\|x_\mu - x_\mu^{\mathcal{N}}\|_{L_2} \le \frac{\gamma_\mu}{\beta_\mu^{\mathcal{N}}} \inf_{v^{\mathcal{N}} \in X_\mu^{\mathcal{N}}} \|x_\mu - v^{\mathcal{N}}\|_{L_2}, \tag{3.2}$$

so that the Petrov-Galerkin approximation is the best approximation (i.e., an *identity*) for $\gamma_\mu = \beta_\mu^{\mathcal{N}} = 1$.

The Petrov-Galerkin framework induces a residual-based error estimation in a straightforward manner. To describe it, let us recall that the *residual* is defined for some $\tilde{x} \in L_2(I)^n$ as

$$r(\tilde{x}) \in Y', \qquad r(\tilde{x})[y] := f_\mu(y) - b_\mu(\tilde{x}, y), \quad y \in Y.$$

Then, it is a standard estimate that

$$\|x_\mu - x_\mu^{\mathcal{N}}\|_{L_2} \le \frac{1}{\beta_\mu} \sup_{y \in Y} \frac{b_\mu(x_\mu - x_\mu^{\mathcal{N}}, y)}{|||y|||_\mu} = \frac{1}{\beta_\mu} |||r(x_\mu^{\mathcal{N}})|||'_\mu =: \Delta_\mu^{\mathcal{N}} \tag{3.3}$$

and $\Delta_\mu^{\mathcal{N}}$ is a residual-based error estimator. Note that for $\beta_\mu = 1$ we have a error-residual identity $\|x_\mu - x_\mu^{\mathcal{N}}\|_{L_2} = |||r(x_\mu^{\mathcal{N}})|||'_\mu = \Delta_\mu^{\mathcal{N}}$.

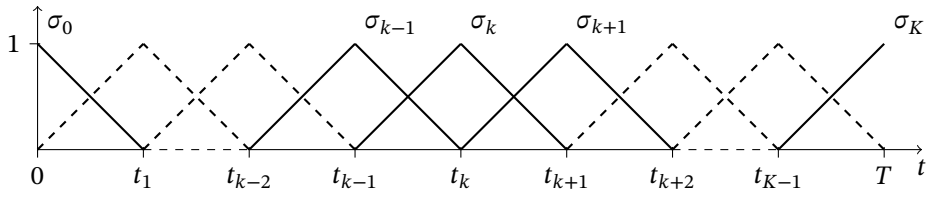

FIGURE 1. Piecewise linear temporal discretization (hat functions).

3.2. **PDAE Petrov-Galerkin Discretization.** We are now going to specify the above general framework to PDAEs. This means that we need to introduce a suitable discretization in time. We fix a constant time step size $\Delta t := T/K$ (i.e., $K \in \mathbb{N}$ is the number of time intervals) and choose for simplicity equidistant nodes $t_k := k\Delta t$, $k = 0, ..., K$ in I. Denote by σ_k, $k = 0, ..., K$ piecewise linear splines corresponding to the nodes t_{k-1}, t_k and t_{k+1}, see Figure 1. For $k \in \{0, K\}$, the hat functions are restricted to the interval $\bar{I}$. For realizing a discretization of higher order, one could simply use splines of higher degree.

As in [6], we start by defining the test space and then construct inf-sup optimal trial spaces. To this end, let $d := \dim(\ker E^T)$ and assume that we have a basis $\{v_1, \dots, v_d\}$ of $\ker E^T$ at hand[5] and form a matrix $V := (v_1, ..., v_d) \in \mathbb{R}^{n\times d}$ by arranging the vectors as columns of V. Then, we construct $Y^{\mathcal{N}} \subset Y = H^1_E(I)^n$ independent of the parameter and choose the trial space as $X^{\mathcal{N}}_\mu := B^*_\mu Y^{\mathcal{N}}$, which will then guarantee that $\beta^{\mathcal{N}}_\mu = 1$. We suggest a piecewise linear discretization by

$$Y^{\mathcal{N}} := \operatorname{span}\{e_i\sigma_k \,:\, k = 0, ..., K-1,\ i = 1, ..., n\} \oplus \operatorname{span}\{v_i\sigma_K \,:\, i = 1, \dots, d\} \subset Y,$$

where $e_i \in \mathbb{R}^n$ denotes i-th canonical vector. Then, we set

$$\begin{aligned} X^{\mathcal{N}}_\mu := B^*_\mu Y^{\mathcal{N}} = \operatorname{span}\{&-E^T e_i\dot{\sigma}_k - A^T_\mu e_i\sigma_k \,:\, k = 0, ..., K-1,\ i = 1, ..., n\} \\ &\oplus \operatorname{span}\{-E^T v_i\dot{\sigma}_K - A^T_\mu v_i\sigma_K \,:\, i = 1, \dots, d\} \subset X = L_2(I)^n, \end{aligned}$$

with dimensions $\mathcal{N} := \dim(X^{\mathcal{N}}) = \dim(Y^{\mathcal{N}}) = nK + d$. Then, Lemma 2.5 and Remark 3.1 ensure $\beta^{\mathcal{N}}_\mu = \beta_\mu = \gamma_\mu = 1$ and thus

$$\|x_\mu - x^{\mathcal{N}}_\mu\|_{L_2} = \inf_{v^{\mathcal{N}} \in X^{\mathcal{N}}_\mu} \|x_\mu - v^{\mathcal{N}}\|_{L_2} = |||r(x^{\mathcal{N}}_\mu)|||'_\mu = \Delta^{\mathcal{N}}_\mu. \tag{3.4}$$

3.2.1. *The linear system.* To construct the discrete linear system $\boldsymbol{B}^{\mathcal{N}}_\mu \boldsymbol{x}^{\mathcal{N}}_\mu = \boldsymbol{f}^{\mathcal{N}}_\mu$ we need bases $\{\xi_1(\mu), \dots, \xi_{\mathcal{N}}(\mu)\}$ of $X^{\mathcal{N}}_\mu$ and $\{\psi_1, \dots, \psi_{\mathcal{N}}\}$ of $Y^{\mathcal{N}}$. The stiffness matrix $\boldsymbol{B}^{\mathcal{N}}_\mu \in \mathbb{R}^{\mathcal{N}\times\mathcal{N}}$ can be computed by $[\boldsymbol{B}^{\mathcal{N}}_\mu]_{j,i} := b_\mu(\xi_i(\mu), \psi_j) = (\xi_i(\mu), B^*_\mu\psi_j)_{L_2}$. We recall that $X^{\mathcal{N}}_\mu = B^*_\mu Y^{\mathcal{N}}$, which implies that $\xi_i(\mu) = B^*_\mu\psi_i$, so that $[\boldsymbol{B}^{\mathcal{N}}_\mu]_{j,i} = (B^*_\mu\psi_i, B^*_\mu\psi_j)_{L_2}$ and $\boldsymbol{B}^{\mathcal{N}}_\mu$ is in fact symmetric positive definite.

The right-hand side $\boldsymbol{f}^{\mathcal{N}}_\mu \in \mathbb{R}^{\mathcal{N}}$ reads $[\boldsymbol{f}^{\mathcal{N}}_\mu]_j := f_\mu(\psi_j)$. The discrete solution then reads $x^{\mathcal{N}}_\mu := \sum_{i=1}^{\mathcal{N}} [\boldsymbol{x}^{\mathcal{N}}_\mu]_i\, \xi_i(\mu)$.

[5]This is in fact the reason why we restricted ourselves to parameter-independent matrices E instead of E_μ. We would then need to have a parameter-dependent basis for $\ker E^T_\mu$, which is of course possible, but causes a quite heavy notation.

Recalling the finite element functions σ_k in Figure 1, we define the inner product matrices for $k, \ell = 0, ..., K$ by

$$[\boldsymbol{K}_{\Delta t}]_{k,\ell} := (\dot{\sigma}_k, \dot{\sigma}_\ell)_{L_2(I)}, \qquad [\boldsymbol{L}_{\Delta t}]_{k,\ell} := (\sigma_k, \sigma_\ell)_{L_2(I)}, \qquad [\boldsymbol{O}_{\Delta t}]_{k,\ell} := (\dot{\sigma}_k, \sigma_\ell)_{L_2(I)},$$

and subdivide the matrices $\boldsymbol{\Pi}_{\Delta t} \in \mathbb{R}^{(K+1)\times(K+1)}$ for $\boldsymbol{\Pi}_{\Delta t} \in \{\boldsymbol{K}_{\Delta t}, \boldsymbol{L}_{\Delta t}, \boldsymbol{O}_{\Delta t}\}$ according to

$$\boldsymbol{\Pi}_{\Delta t} = \begin{pmatrix} \boldsymbol{\Pi}_{\Delta t}^{1,1} & \boldsymbol{\Pi}_{\Delta t}^{1,2} \\ \boldsymbol{\Pi}_{\Delta t}^{2,1} & \boldsymbol{\Pi}_{\Delta t}^{2,2} \end{pmatrix}, \quad \begin{matrix} \boldsymbol{\Pi}_{\Delta t}^{1,1} \in \mathbb{R}^{K\times K}, & \boldsymbol{\Pi}_{\Delta t}^{1,2} \in \mathbb{R}^{K\times 1}, \\ \boldsymbol{\Pi}_{\Delta t}^{2,1} \in \mathbb{R}^{1\times K}, & \boldsymbol{\Pi}_{\Delta t}^{2,2} \in \mathbb{R}. \end{matrix}$$

Then, the stiffness matrix also has block structure

$$\boldsymbol{B}_\mu^{\mathcal{N}} = \begin{pmatrix} \boldsymbol{B}_\mu^{1,1} & \boldsymbol{B}_\mu^{1,2} \\ \boldsymbol{B}_\mu^{2,1} & \boldsymbol{B}_\mu^{2,2} \end{pmatrix} \in \mathbb{R}^{\mathcal{N}\times\mathcal{N}}$$

in form of Kronecker products of matrices, i.e. (with $V = (v_1, \dots, v_d) \in \mathbb{R}^{n\times d}$ as above),

$$\begin{aligned}
\boldsymbol{B}_\mu^{1,1} &= \boldsymbol{K}_{\Delta t}^{1,1} \otimes EE^T + \boldsymbol{O}_{\Delta t}^{1,1} \otimes EA_\mu^T + (\boldsymbol{O}_{\Delta t}^{1,1})^T \otimes A_\mu E^T + \boldsymbol{L}_{\Delta t}^{1,1} \otimes A_\mu A_\mu^T \in \mathbb{R}^{nK\times nK}, \\
\boldsymbol{B}_\mu^{1,2} &= \boldsymbol{O}_{\Delta t}^{1,2} \otimes EA_\mu^T V + \boldsymbol{L}_{\Delta t}^{1,2} \otimes A_\mu A_\mu^T V \in \mathbb{R}^{nK\times d}, \\
\boldsymbol{B}_\mu^{2,1} &= (\boldsymbol{O}_{\Delta t}^{1,2})^T \otimes V^T A_\mu E^T + \boldsymbol{L}_{\Delta t}^{2,1} \otimes V^T A_\mu A_\mu^T \in \mathbb{R}^{d\times nK}, \\
\boldsymbol{B}_\mu^{2,2} &= \boldsymbol{L}_{\Delta t}^{2,2} \otimes V^T A_\mu A_\mu^T V \in \mathbb{R}^{d\times d}.
\end{aligned}$$

For the right-hand side, given some function $f_\mu : \bar{I} \to \mathbb{R}^n$, we obtain a discretization $\boldsymbol{f}_\mu^{\mathcal{N}} \in \mathbb{R}^{\mathcal{N}}$ in the sense of (2.7) by $[\boldsymbol{f}_\mu^{\mathcal{N}}]_j = \sum_{k=0}^K (f_\mu(t_k)\sigma_k, \psi_j)_{L_2}$, $j = 1, ..., \mathcal{N}$. This means that we discretize f_μ in time by means of piecewise linears. Collecting the sample values of f_μ in one vector, i.e., $\boldsymbol{f}_{\mu,\Delta t} := \big(f_\mu(t_0), \dots, f_\mu(t_K)\big)^T \in \mathbb{R}^{n(K+1)}$ we get that

$$\boldsymbol{f}_\mu^{\mathcal{N}} = \boldsymbol{F}_{\Delta t}^T \boldsymbol{f}_{\mu,\Delta t} \quad \text{where} \quad \boldsymbol{F}_{\Delta t} := \begin{pmatrix} Id_n \otimes \boldsymbol{L}_{\Delta t}^{1,1} & V \otimes \boldsymbol{L}_{\Delta t}^{1,2} \\ Id_n \otimes \boldsymbol{L}_{\Delta t}^{2,1} & V \otimes \boldsymbol{L}_{\Delta t}^{2,2} \end{pmatrix} \in \mathbb{R}^{n(K+1)\times\mathcal{N}}$$

and $Id_n \in \mathbb{R}^{n\times n}$ again denoting the n-dimensional identity matrix.

As already noted above, of course, one could use different discretizations (e.g. higher order or different discretizations for f_μ and the test functions) and we choose the described one just for simplicity.

The efficient numerical solution of this linear system requires a solver that takes the specific structure into account. For similar systems arising from space-time (ultraweak) variational formulations of heat, transport and wave equations, such specific efficient solvers have been introduced in [15, 16]. The structure of the system above is different and we will consider the development of efficient solvers in future research, see Section 6.

3.2.2. *Special case: Linear DAEs.* We are going to specify the above general setting to the special case of fully linear DAEs, namely

$$E\dot{x}(t) - Ax(t) = Bu(t) + g_{\mu_2}(t), \quad t \in I, \qquad x(0) = 0, \tag{3.5}$$

in which the right-hand side is given in terms of a matrix $B \in \mathbb{R}^{n\times m}$, a control $u : \bar{I} \to \mathbb{R}^m$, m denoting some input dimension and a function $g_{\mu_2} : \bar{I} \to \mathbb{R}^n$, which arises from the reduction to homogeneous initial conditions, see §1.3. The initial condition

is assumed to be parameterized through g_{μ_2} by $\mu_2 \in \mathcal{P}_2 \subset \mathbb{R}^{P_2}$, $P_2 \in \mathbb{N}$. In view of (1.2) and §1.3, we get

$$g_{\mu_2}(t) = \sum_{q=1}^{Q_x} \vartheta_q^x(\mu_2)\left(A\bar{x}_q(t) - E\dot{\bar{x}}_q(t)\right) =: \sum_{q=1}^{Q_x} \vartheta_q^x(\mu_2)\, z_q(t),$$

where $\bar{x}_q \in C^1(\bar{I})^n$ are smooth extensions of $\tilde{x}_{0,q}$, i.e., $\bar{x}_q(0) = \tilde{x}_{0,q}$, $q = 1, ..., Q_x$.

We view the control and the initial condition (via g_{μ_2}) as parameters, i.e., $f_\mu(t) = B\mu_1(t) + g_{\mu_2}(t)$, $\mu = (\mu_1, \mu_2)$, which means that the parameter set would be infinite-dimensional and needs to be discretized. Using the same kind of discretization as above, we can use the samples of the control as parameter, i.e.,

$$\boldsymbol{\mu}_1 := \left(u(t_0), ..., u(t_K)\right)^T \in \mathcal{P}_1 = \mathbb{R}^{P_1}, \quad P_1 = m(K+1),$$

and similar for the initial condition $\boldsymbol{z}_q := \left(z_q(t_0), ..., z_q(t_K)\right)^T \in \mathbb{R}^{n(K+1)}$, $q = 1, ..., Q_x$. Then, we get

$$\boldsymbol{f}_\mu^{\mathcal{N}} = \boldsymbol{F}_{\Delta t}^T (B \otimes Id_{K+1})\, \boldsymbol{\mu}_1 + \sum_{q=1}^{Q_x} \vartheta_q^x(\mu_2)\, \boldsymbol{F}_{\Delta t}^T \boldsymbol{z}_q,$$

so that the parameter dimension is $P = P_1 + P_2 = m(K+1) + P_2$, which might be large. The right-hand side $\boldsymbol{f}_\mu^{\mathcal{N}}$ thus also admits an affine decomposition with $Q_f = P_1 + Q_x = m(K+1) + Q_x$ terms. However, this number might be an issue concerning efficiency if K is large. Nevertheless, if $m \ll n$, we still have $Q_f \ll \mathcal{N}$.

Moreover, in the linear case, the matrix $A_\mu \equiv A$ is independent of the parameter, which means (among other facts) that trial and test spaces are parameter-independent, as sets and also w.r.t. their topology. Note that this is the most common case for system theoretic MOR methods (like BT), which are often even restricted to this case, [20, 25], with the exception [8]. Our setting seems more flexible in this regard and fully linear DAEs are just a special case.

4. Model order reduction: The Reduced Basis Method

The Reduced Basis Method (RBM) is a model order reduction technique which has originally been constructed for parameterized partial differential equations (PPDEs), see e.g. [3, 17, 22]. In an offline training phase, a reduced basis of size $N \ll \mathcal{N}$ is constructed (typically in a greedy manner, see Algorithm 1 below) from sufficiently detailed approximations for certain parameter samples (also called "truth" approximations or *snapshots*), which are computed e.g. by a Petrov-Galerkin method as described above. In particular, $\mathcal{N}$ is assumed to be sufficiently large in order to ensure that $x_\mu^{\mathcal{N}}$ is (at least numerically) indistinguishable from the exact state x_μ, which explains the name "truth". As long as an efficient solver for the detailed problem is available, we may assume that the snapshots can be computed in $\mathcal{O}(\mathcal{N})$ complexity.

Given some parameter value μ, the reduced approximation $x_N(\mu)$[6] is then computed by solving a reduced system of dimension N. Thanks to the affine decomposition (1.2), several quantities for the reduced system can be precomputed and stored so that a reduced approximation is determined in $\mathcal{O}(N^3)$ operations, independent of $\mathcal{N}$ (which is called *online efficient*). Moreover, an *a posteriori* error estimator $\Delta_N(\mu)$ guarantees a

[6]For all quantities of the reduced system, we write the parameter μ as an argument in order to clearly distinguish the detailed approximation $x_\mu^{\mathcal{N}}$ from the reduced approximation $x_N(\mu)$ for the same parameter.

certification in terms of an online efficiently computable upper bound for the error, i.e., $\|x_\mu^{\mathcal{N}} - x_N(\mu)\|_{L_2} \leq \Delta_N(\mu)$.

We are going to use this framework for PDAEs of the form (1.1). Model reduction of (1.1) may be concerned (at least) with the following quantities

- size n of the system,
- dimension K of the temporal discretization,

where we have in mind to solve (1.1) extremely fast for several values of the parameter μ. As mentioned earlier, the first issue has extensively been studied in the literature e.g. by system theoretic methods, in particular for fully linear DAEs (3.5). This can be done independently from the subsequent reduction w.r.t. K (both for parameterized and non-parameterized versions), so that we even might assume that such Model Order Reduction (MOR) techniques have already been applied in a preprocessing step. We mention [8] for a system theoretic MOR for parameter-dependent DAEs. Here, we are going to consider the reduction w.r.t. time using the RBM based upon a variational formulation w.r.t. the time variable.

We restrict ourselves to the reduction of the fully linear case of (3.5) as it easily shows how the RBM-inspired model reduction can be combined with existing system theoretic approaches to reduce the size of the system (e.g. in a preprocessing step). In the fully linear case, the matrix $A_\mu \equiv A$ and hence all operators and bilinear forms on the left-hand side are parameter-independent. This implies in addition that the ansatz space $X_\mu^{\mathcal{N}} \equiv X^{\mathcal{N}}$ and the norm $|||\cdot|||_\mu \equiv |||\cdot|||$ inducing the topology on the test space are parameter-independent as well, which of course simplifies the framework. However, parameter-dependent matrices A_μ can be treated similar to the RBM for ultraweak formulations of PPDEs as described e.g. in [6, 16, 30]. However, we note that the RB approach also allows the treatment of more general PDAEs and is not restricted to fully linear systems (3.5), in particular w.r.t. the right-hand side.

The idea of the RBM can be described as follows: One determines sample values

$$S_N := \{\mu^{(1)}, ..., \mu^{(N)}\} \subset \mathcal{P}$$

of the parameters in an offline training phase by a greedy procedure described in Algorithm 1 below. Then, for each $\mu \in S_N$, we determine a sufficiently detailed "snapshot" $x_\mu^{\mathcal{N}} \in X^{\mathcal{N}}$ by the ultraweak Petrov-Galerkin discretization as in §3.2 and obtain a reduced space of dimension N by setting

$$X_N := \operatorname{span}\{x_\mu^{\mathcal{N}} : \mu \in S_N\} =: \operatorname{span}\{\zeta_1, ..., \zeta_N\} \subset X^{\mathcal{N}}.$$

We also need a reduced *test* space for the Petrov-Galerkin method. Recalling that the operator is parameter-independent here ($B_\mu \equiv B$) and also the trial space $X^{\mathcal{N}}$ is independent of μ, we can easily identify the optimal test space. In fact, for each snapshot there exists a unique $y_\mu^{\mathcal{N}} \in Y^{\mathcal{N}}$ such that $x_\mu^{\mathcal{N}} = B^* y_\mu^{\mathcal{N}}$. Then, we define

$$Y_N := \operatorname{span}\{y_\mu^{\mathcal{N}} : \mu \in S_N\} =: \operatorname{span}\{\eta_1, ..., \eta_N\} \subset Y^{\mathcal{N}}.^7$$

Then, given a new parameter value $\mu \in \mathcal{P}$, one determines the *reduced approximation* $x_N(\mu) \in X_N$ by solving (recall that here $b_\mu \equiv b$)

$$b(x_N(\mu), y_N) = f_\mu(y_N) \quad \text{for all } y_N \in Y_N.$$

If $N \ll \mathcal{N} = nK + d$, we can compute a reduced approximation with significantly less effort as compared to the Petrov-Galerkin (or a time-stepping) method. To determine the reduced approximation $x_N(\mu)$, we have to solve a linear system of the form

[7] For efficiency reasons, in fact, we first determine η_i and then simply set $\zeta_i := B^* \eta_i$.

$\boldsymbol{B}_N \boldsymbol{x}_N(\mu) = \boldsymbol{f}_N(\mu)$, where the stiffness matrix is given by $[\boldsymbol{B}_N]_{j,i} = b(\zeta_i, \eta_j)$, $i, j = 1, ..., N$, recalling that the bilinear form is parameter-independent. Hence, $\boldsymbol{B}_N \in \mathbb{R}^{N\times N}$ can be computed and stored in the offline phase. For the right-hand side, we use the affine decomposition (1.2) and get $[\boldsymbol{f}_N(\mu)]_j = \sum_{q=1}^{Q_f} \vartheta_g^f(\mu)\,(\tilde{f}_q, \eta_j)_{L_2}$. The quantities $(\tilde{f}_q, \eta_j)_{L_2}$ can be precomputed and stored in the offline phase, so that $\boldsymbol{f}_N(\mu)$ is computed online efficient in $\mathcal{O}(Q_f N)$ operations. Obtaining the coefficient vector $\boldsymbol{x}_N(\mu)$, the reduced approximation results in $x_N(\mu) = \sum_{i=1}^{N} [\boldsymbol{x}_N(\mu)]_i\, \zeta_i$. Note that the matrix $\boldsymbol{B}_N$ is typically densely populated so that the numerical solution requires in general $\mathcal{O}(N^3)$ operations.

The announced greedy selection of the samples is based upon the residual error estimate (here identity) $\Delta_\mu^{\mathcal{N}}$ in (3.3) resp. (3.4) for the reduced system described as follows: In a similar manner as deriving $\Delta_\mu^{\mathcal{N}}$ in (3.3) we get a residual based error estimator for the reduced approximation

$$\|x_\mu^{\mathcal{N}} - x_N(\mu)\|_{L_2} \leq \frac{1}{\beta^{\mathcal{N}}} \sup_{y\in Y^{\mathcal{N}}} \frac{b(x_\mu^{\mathcal{N}} - x_N(\mu), y)}{|||y|||} = \frac{1}{\beta^{\mathcal{N}}} |||r(x_N(\mu))|||' =: \Delta_N(\mu),$$

since the bilinear form and the norm in Y are parameter-independent here. Hence, the inf-sup constant is parameter-independent as well, i.e., $\beta_\mu^{\mathcal{N}} \equiv \beta^{\mathcal{N}}$ and it is unity by Remark 3.1, so that

$$\|x_\mu^{\mathcal{N}} - x_N(\mu)\|_{L_2} = |||r(x_N(\mu))|||' = \Delta_N(\mu). \tag{4.1}$$

Its computation can be done in an online efficient manner in $\mathcal{O}(N)$ operations by determining Riesz representations in the offline phase, see [3, 17, 22]. We use this error identity in the greedy method in Algorithm 1 below.

Algorithm 1 (Weak) Greedy method

input: training sample $\mathcal{P}_{\text{train}} \subseteq \mathcal{P}$, tolerance $\varepsilon > 0$, max. dimension $N_{\max} \in \mathbb{N}$

choose $\mu^{(1)} \in \mathcal{P}_{\text{train}}$, compute snapshot $\zeta_1 := x_{\mu^{(1)}}^{\mathcal{N}}$ and optimal test function η_1 with $\zeta_1 = B^*\eta_1$

Initialize $S_1 \leftarrow \{\mu^{(1)}\}$, $X_1 := \text{span}\{\zeta_1\}$, $Y_1 := \text{span}\{\eta_1\}$, $N := 1$

while $N < N_{\max}$ **do**

  if $\max_{\mu\in\mathcal{P}_{\text{train}}} \Delta_N(\mu) \leq \varepsilon$ **then return**

  $\mu^{(N+1)} \leftarrow \arg\max_{\mu\in\mathcal{P}_{\text{train}}} \Delta_N(\mu)$

  compute snapshot $\zeta_{N+1} := x_{\mu^{(N+1)}}^{\mathcal{N}}$ and optimal test function η_{N+1}

  $S_{N+1} \leftarrow S_N \cup \{\mu^{(N+1)}\}$, $X_{N+1} := X_N \oplus \text{span}\{\zeta_{N+1}\}$, $Y_{N+1} := Y_N \oplus \text{span}\{\eta_{N+1}\}$

  $N \leftarrow N + 1$

end while

output: set of chosen parameters S_N, reduced spaces X_N, Y_N

5. Numerical experiments

In this section, we report on results of some of our numerical experiments. Our main focus is on the numerical solution of the ultraweak form of the PDAE, the error estimation and the quantitative reduction. We solve the arising linear systems for the

Petrov-Galerkin and the reduced system by MATLAB's backslash operator, see also our remarks in Section 6 below. The codes for producing the subsequent results is available via `https://github.com/mfeuerle/Ultraweak_PDAE`.

5.1. **Serial RLC circuit.** We start by a standard problem which (in some cases) admits a closed formula for the analytical solution. This allows us to monitor the exact error and a comparison with standard time-stepping methods. Our particular interest is the approximation property of the ultraweak approach, which is an L_2-approximation.

The serial RLC circuit consists of a resistor with resistance R, an inductor with inductance L, a capacitor with capacity C and a voltage source fed by a voltage curve $f_{V_S} : \bar{I} \to \mathbb{R}$. Kirchhoff's circuit and further laws from electrical engineering yield a DAE with the data

$$E = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \; A = \begin{pmatrix} 0 & 0 & L^{-1} & 0 \\ C^{-1} & 0 & 0 & 0 \\ R & 0 & 0 & -1 \\ 0 & 1 & 1 & 1 \end{pmatrix}, \; x = \begin{pmatrix} x_I \\ x_{V_C} \\ x_{V_L} \\ x_{V_R} \end{pmatrix}, \; f = \begin{pmatrix} 0 \\ 0 \\ 0 \\ -f_{V_S} \end{pmatrix},$$

whose index is $k = 1$. The solution x consists of the electric current x_I and the voltages at the capacitor x_{V_C}, at the inductor x_{V_L} and at the resistor x_{V_R}.

Convergence of the Petrov-Galerkin scheme. In Figure 2, we compare the exact solution with approximations generated by a standard time-stepping scheme (using MATLAB's fully implicit variable order solver with adaptive step size control *ode15i*, [19]) and by our ultraweak formulation from §3.2. We choose two specific examples for f_{V_S}, namely a smooth and a discontinuous one,

$$f_{V_S}^{\text{smooth}}(t) := \sin\big(\frac{4\pi}{T}t\big), \qquad f_{V_S}^{\text{disc}}(t) := \operatorname{sign}\Big(\cos\big(\frac{4\pi}{T}t\big)\Big).$$

For the smooth right-hand side (left graph in Figure 2), both *ode15i* and the ultraweak method give good results. Concerning the deviations for the ultraweak approach at the start and end time, we recall that the ultraweak form yields an approximation in L_2, so that pointwise comparisons are not necessarily meaningful.

In the discontinuous case, existence of a classical solution cannot be guaranteed by the above arguments. In particular, there is no closed solution formula. As we see in the right graph in Figure 2, *ode15i* stops at the first jump. This is to be expected, since $f_{V_S}^{\text{disc}} \notin C^0(\bar{I})$, so that the solution lacks sufficient regularity to guarantee convergence of a time-stepping scheme like *ode15i* (even though it is an adaptive variable order method). We could resolve the jumps even better by choosing more time steps K, while *ode15i* still fails. We conclude that the ultraweak method also converges for problems lacking regularity.

Convergence rate. Next, we investigate the rate of convergence for the ultraweak form. To that end, we use $f_{V_S}^{\text{smooth}}$, since the analytical solution x^* is known and we can thus compute the relative error $\|x^* - x^{\mathcal{N}}\|_{L_2} / \|x^*\|_{L_2}$. Using the lowest order discretization as mentioned above (namely piecewise linear test functions ψ_j, which yield discontinuous trial functions $B^*\psi_i$), we can only hope for first order (w.r.t. the number of time steps K), which we see in Figure 3 and was observed in all cases we considered. We obtain higher order convergence by choosing test functions of higher order, provided the solution has sufficient smoothness.

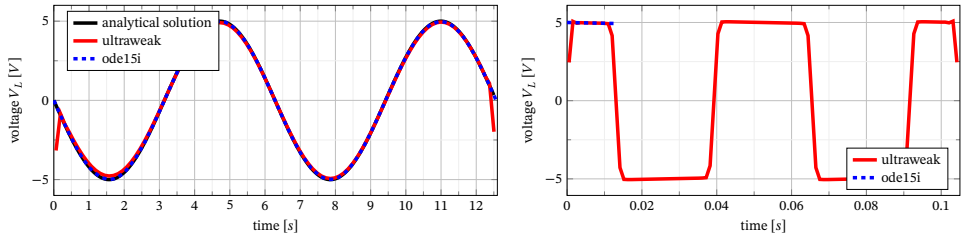

FIGURE 2. Serial RLC circuit, exact voltage at the inductor; comparison of time-stepping (ode15i – blue) and ultraweak (red) approximation for smooth $f_{V_S}^{\text{smooth}}$ (left, including analytical solution) and discontinuous $f_{V_S}^{\text{disc}}$ (right) right-hand side.

Moreover, we compare the exact relative error with our error estimator (see §3.1). Figure 3 shows a perfect matching confirming the error-residual identity (3.4) also for the numerically computed error estimator.

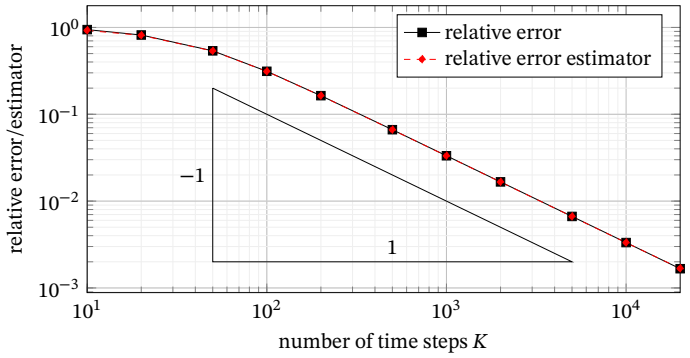

FIGURE 3. Relative error $\|x^* - x^{\mathcal{N}}\|_{L_2}/\|x^*\|_{L_2}$ and relative error estimator $\Delta^{\mathcal{N}}/\|x^*\|_{L_2}$ w.r.t. to the analytical solution x^* for increasing numbers of time steps K.

5.2. **Time dependent Stokes problem.** In order to investigate the quantitative performance of the model reduction, we consider a problem, which has often been used as a benchmark, [21, 23, 27, 28], namely the time-dependent Stokes problem on the unit square $(0,1)^2$, discretized by a finite volume method on a uniform, staggered grid for the spatial variables with n unknowns, [28], where we choose $n = 644$. The arising homogeneous fully linear DAE with output function $y : I \to \mathbb{R}$ takes the form (3.5),

$$E\dot{x}(t) - Ax(t) = Bu(t) + g(t), \quad t \in I, \qquad x(0) = 0, \tag{5.1a}$$

$$y(t) = Cx(t), \tag{5.1b}$$

where $C \in \mathbb{R}^{1\times n}$ is an output matrix, $B \in \mathbb{R}^{n\times 1}$ is the control matrix, and $u : I \to \mathbb{R}$ is a control, which serves as a parameter $\mu \equiv u$ as described in §3.2.2 above.[8] We use a parameter-independent initial condition, so that $g_\mu \equiv g$ and $Q_x = 1$.

[8]We could also choose larger input/output-dimensions.

In order to combine system theoretic model reduction with the Reduced Basis Method from §4, we use the system theoretic model order reduction package [21]. In particular, we use Balanced Truncation (BT) from [27] during a preprocessing step to reduce the above system of dimension n to a system

$$\hat{E}\dot{\hat{x}}(t) - \hat{A}\hat{x}(t) = \hat{B}u(t) + \hat{g}(t), \quad t \in I, \qquad \hat{x}(0) = 0, \tag{5.2a}$$

$$y(t) = \hat{C}\hat{x}(t), \tag{5.2b}$$

with $\hat{E}, \hat{A} \in \mathbb{R}^{\hat{n}\times\hat{n}}$, $\hat{B} \in \mathbb{R}^{\hat{n}\times 1}$, $\hat{C} \in \mathbb{R}^{1\times\hat{n}}$ as well as $\hat{x}, \hat{g} : \bar{I} \to \mathbb{R}^{\hat{n}}$ and $\hat{n} \ll n$. We note that the resulting reduced system typically provides regular matrices $\hat{E}, \hat{A}$. Then, the reduced system is a linear time-invariant system (LTI), which is an easier problem than a DAE and in fact a special case. Hence, our presented approach is still valid, even though designed for PDAEs. For an ultraweak formulation of LTI systems, we refer to [10].

Remark 5.1. We use the RBM here for deriving a certified reduced approximation of the state x. If we would want to control the output y along with a corresponding error estimator Δ_N^y, it is fairly standard in the theory of RBM to use a primal-dual approach with a second (dual) reduced basis, e.g. [17, 22]. For simplicity of exposition, we leave this to future work and compute the output from the state by $\hat{C}\hat{x}(t)$, resp. $Cx(t)$. ⋄

Discretization of the control within the RBM. Since we use a variational approach, we are in principle free to choose any discretization for the control (we only need to compute inner products with the test basis functions). We tested piecewise linear discretizations as described in §3.2 for different step sizes K_u/T, where K_u might be different from K, which we choose for discretizing the state. Doing so, the parameter reads $\mu = (u(t_0), \dots, u(t_{K_u}))^T \in \mathcal{P} \equiv \mathbb{R}^{K_u+1}$, i.e., the parameter dimension is $P = K_u + 1$, which might be large. Large parameter dimensions are potentially an issue for the RBM since the curse of dimension occurs. Hence, we investigate if we can reduce K_u within the RBM.

In order to answer this question, we apply Algorithm 1 to the time-dependent Stokes problem (5.1a) (without BT) setting $\varepsilon = 0$, $N_{\max} = Q_f$ from (1.2) (i.e., $Q_f = P + 1$ for the fully linear system with parameter-independent initial value) and P_{train} consisting of 500 random vectors for $K_u \equiv K \in \{75, 150, 300\}$, i.e., $\mathcal{N} = 48\,524, 96\,824, 193\,424$, where $d = 224$. For these three cases, we investigate the *max greedy training error*, i.e., $\max_{\mu\in\mathcal{P}_{\text{train}}} \Delta_N(\mu)$. The results in Figure 4 show an exponential decay w.r.t. the dimension N of the reduced system with slower decay as K grows. This is to be expected as the discretized control space is much richer for growing K_u and the reduced model has to be able to represent this variety. However, in relative terms (i.e., reduced size N compared with full size K), we see that the compression rates are almost the same. This shows that the RBM can effectively reduce the system no matter how strong the influence of the control on the state is. It is expected that this potential is even more pronounced if a primal-dual RBM is used for the output.

Next, we note that for $A \equiv A_\mu$ as (5.1a), the reduced model is always *exact* for $N \geq Q_f$, which explains the drop off of the curves in Figure 4. For fully linear DAEs, a reduced model with $N \geq Q_f = K_u + 2$ is always exact. Hence, if $m \ll n$ (here $m = 1 \ll 644 = n$), we obtain an exact reduced model of dimension $N = Q_f = P + Q_x = m(K_u + 1) + 1 \ll nK + d =: \mathcal{N}$. Even though this seems to be attractive for low-dimensional outputs, we stress the fact that the reduced dimension still depends on

the temporal dimension K_u, which might be large. Hence, a combination of a possibly small discretization of the control and a RBM seems necessary.

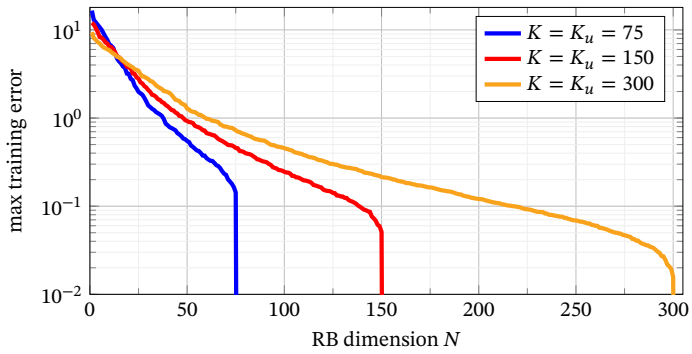

FIGURE 4. Maximal greedy training error $\max_{\mu\in\mathcal{P}_{\text{train}}} \Delta_N(\mu)$ for different time resolutions $K_u = K \in \{75, 150, 300\}$ over the reduced dimension N.

Let us comment on the error decay of the RBM produced by the greedy method using the error estimator derived from the ultraweak formulation of the PDAE. We obtain exponential decay of the error, which in fact shows the potential of the RBM. The question if a given PDAE permits a fast decay of the greedy RBM error is well-known to be linked to the decay of the Kolmogorov N-width, [3, 17, 30], which is a property only of the problem at hand. In other words, if a PDAE can be reduced w.r.t. time, the greedy method will detect this.

The results in Figure 4 use $K_u = K$. The next question is how the error behaves for $K_u < K$. To this end, we determine the error in the state w.r.t. the full resolution, i.e., we compare the state derived from the control with K_u degrees of freedom with the state of the fully resolved control. In Figure 5, we display errors for different values for K. We obtain fast convergence, which again shows the significant potential for a reduced temporal discretization of the control.

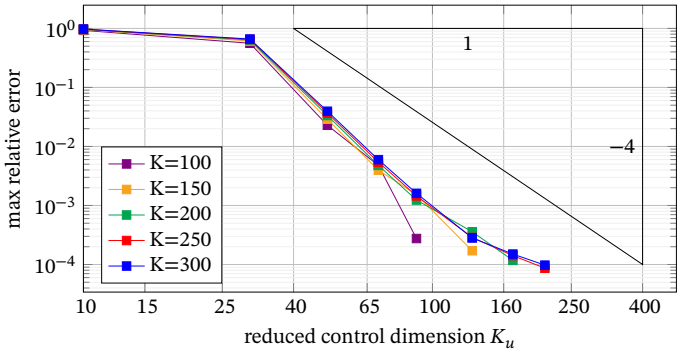

FIGURE 5. Max error for control dimensions of size $K_u < K$.

Combination with BT / RBM error decay. Next, we wish to investigate if a combination of a system theoretic MOR (here BT) and an RBM-like reduction w.r.t. time can be combined. To this end, we fix the temporal resolution (i.e., the number of time steps, here $K = K_u = 100$) and determine the RBM error using Algorithm 1 for the full and the BT-reduced system. We use [21] to compute the BT from [27] and obtain a LTI system of dimension $\hat{n} = 5$.

The results are shown in Figure 6, where we again show the maximal training error. As we see, the error for the BT-reduced system is smaller than the original one, which in fact indicates that we can combine both methods. We got similar results for other choices of K. This shows that there is as much "reduction potential" in the reduced system (5.2a) as in the original system (5.1a). In other words, a combination of BT and RBM shows significant compression potential.

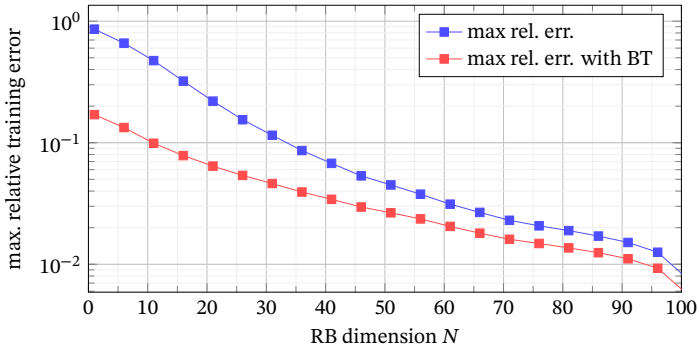

FIGURE 6. Maximal RBM relative error decay over the reduced dimension N for the full system (5.1a) (blue) and for the reduced system (5.2a) with $K = 100$ (red).

6. CONCLUSIONS AND OUTLOOK

In this paper, we introduced a well-posed ultraweak formulation for DAEs and an optimally stable Petrov-Galerkin discretization, which admits a sharp error bound. The scheme shows the expected order of convergence depending on the regularity of the solution and the smoothness of the trial functions. The scheme also converges in low-regularity cases, where classical standard time-stepping schemes fail. Moreover, the stability of the Petrov-Galerkin scheme allows us to choose *any* temporal discretization without satisfying other stability criteria like a CFL condition.

Based upon the ultraweak framework, we introduced a model order reduction in terms of the Reduced Basis Method with an error/residual identity. We have obtained fast convergence and the possibility to combine the RBM for a reduction w.r.t. time with system theoretic methods such as Balanced Truncation to reduce the size of the system.

There are several open issues for future research. We already mentioned a primal-dual RBM for an efficient reduction of the output, the generalization to parameter-dependent matrices A_μ and more general DAEs (not only fully linear). We also mentioned that the system matrix is a sum of Kronecker products of high dimension, which

calls for specific solvers as in [16] for the (parameterized) wave equation. Another issue in that direction is the need for a basis of $\ker(E^T)$, which might be an issue for high-dimensional problems.

References

[1] W. Arendt and K. Urban. *Partial Differential Equations: An analytic and numerical approach*. Translated by J.B. Kennedy, to appear. Springer, 2022.

[2] M. Barrault, Y. Maday, N. Nguyen, and A. Patera. "An empirical interpolation method: application to efficient reduced-basis discretization of partial differential equations". *C. R. Acad. Sci. Paris, Ser. I* 339.9 (2004), pp. 667–672.

[3] P. Benner, A. Cohen, M. Ohlberger, and K. Willcox. *Model reduction and approximation: theory and algorithms*. Vol. 15. SIAM, 2017.

[4] P. Benner and T. Stykel. "Model Order Reduction for Differential-Algebraic Equations: A Survey". *Surveys in Differential-Algebraic Equations IV*. Ed. by A. Ilchmann and T. Reis. Springer, 2017, pp. 107–160.

[5] K. E. Brenan, S. L. Campbell, and L. R. Petzold. *Numerical solution of initial-value problems in differential-algebraic equations*. SIAM, 1995.

[6] J. Brunken, K. Smetana, and K. Urban. "(Parametrized) First Order Transport Equations: Realization of Optimally Stable Petrov-Galerkin Methods". *SIAM J. Sci. Comput.* 41.1 (2019), A592–A621.

[7] S. Chaturantabut and D. Sorensen. "Nonlinear Model Reduction via Discrete Empirical Interpolation". *SIAM J. Sci. Comput.* 32.5 (2010), pp. 2737–2764.

[8] S. Chellappa, L. Feng, and P. Benner. "Adaptive basis construction and improved error estimation for parametric nonlinear dynamical systems". *Internat. J. Numer. Methods Engrg.* 121.23 (2020), pp. 5320–5349.

[9] W. Dahmen, C. Huang, C. Schwab, and G. Welper. "Adaptive Petrov-Galerkin methods for first order transport equations". *SIAM J. Numer. Anal.* 50.5 (2012), pp. 2420–2445.

[10] M. Feuerle and K. Urban. "A Variational Formulation for LTI-Systems and Model Reduction". *ENUMATH 2019: Numerical Mathematics and Advanced Applications*. Ed. by F. J. Vermolen and C. Vuik. Springer, 2021, pp. 1059–1067.

[11] F. Gantmacher. *Theory of Matrices II*. Chelsea Publ., 1959.

[12] C. Gear and L. Petzold. "Differential/algebraic systems and matrix pencils". *Matrix pencils*. Ed. by B. Kågström and A. Ruhe. Springer, 1983, pp. 75–89.

[13] S. Grundel, T. Reis, and S. Schöps. *Progress in Differential-Algebraic Equations II*. Springer, 2020.

[14] E. Hairer and G. Wanner. *Solving ordinary differential equations. II*. Vol. 14. Springer, 2010, pp. xvi+614.

[15] J. Henning, D. Palitta, V. Simoncini, and K. Urban. "Matrix Oriented Reduction of Space-Time Petrov-Galerkin Variational Problems". *Numerical Mathematics and Advanced Applications ENUMATH 2019*. Ed. by F. J. Vermolen and C. Vuik. Springer, 2019, pp. 1049–1057.

[16] J. Henning, D. Palitta, V. Simoncini, and K. Urban. *An Ultraweak Space-Time Variational Formulation for the Wave Equation: Analysis and Efficient Numerical Solution*. 2021. arXiv: `2107.12119 [math.NA]`.

[17] J. S. Hesthaven, G. Rozza, and B. Stamm. *Certified Reduced Basis Methods for Parametrized Partial Differential Equations*. Springer, 2016.

[18] P. Kunkel and V. Mehrmann. *Differential-algebraic equations*. Analysis and numerical solution. European Mathematical Society (EMS), Zürich, 2006.
[19] *MATLAB Version 9.11.0 (R2021b)*. The Mathworks, Inc. Natick, Massachusetts, United States, 2021.
[20] V. Mehrmann and T. Stykel. "Balanced Truncation Model Reduction for Large-Scale Systems in Descriptor Form". *Dimension Reduction of Large-Scale Systems*. Ed. by P. Benner, D. C. Sorensen, and V. Mehrmann. Springer, 2005, pp. 83–115.
[21] J. Saak, M. Köhler, and P. Benner. *M-M.E.S.S.-2.2 – The Matrix Equations Sparse Solvers library*. `https://www.mpi-magdeburg.mpg.de/projects/mess`. 2022.
[22] A. Quarteroni, A. Manzoni, and F. Negri. *Reduced basis methods for partial differential equations: an introduction*. Vol. 92. Springer, 2015.
[23] M. Schmidt. "Systematic discretization of input/output maps and other contributions to the control of distributed parameter systems". PhD thesis. TU Berlin, 2007.
[24] S. Schöps, A. Bartel, M. Günther, E. ter Maten, and P. Müller. *Progress in Differential-Algebraic Equations*. Springer, 2014.
[25] T. Stykel. "Balanced truncation model reduction for descriptor systems". *PAMM* 3 (2003), pp. 5–8.
[26] T. Stykel. "Gramian-Based Model Reduction for Descriptor Systems". *Mathematics of Control, Signals and Systems* 16.4 (2004), pp. 297–319.
[27] T. Stykel. "Balanced truncation model reduction for semidiscretized Stokes equation". *Linear Algebra Appl.* 415.2 (2006), pp. 262–289.
[28] The MORwiki Community. *Stokes equation*. MORwiki – Model Order Reduction Wiki. 2018.
[29] S. Trenn. "Solution Concepts for Linear DAEs: A Survey". *Surveys in Differential-Algebraic Equations I*. Ed. by A. Ilchmann and T. Reis. Springer, 2013, pp. 137–172.
[30] K. Urban. *The Reduced Basis Method in Space and Time: Challenges, Limits and Perspectives*. preprint, Ulm University, to appear in CIME lecture notes, Springer. 2022.
[31] J. Xu and L. Zikatanov. "Some observations on Babuška and Brezzi theories". *Numer. Math.* 94.1 (2003), pp. 195–202.

Ulm University, Institute for Numerical Mathematics, Helmholtzstr. 20, D-89081 Ulm, Germany
Email address: {emil.beurer,moritz.feuerle,niklas.reich,karsten.urban}@uni-ulm.de