



# An Ultraweak Variational Method for Parameterized Linear Differential-Algebraic Equations

Emil Beurer, Moritz Feuerle, Niklas Reich and Karsten Urban\*

Institute for Numerical Mathematics, Ulm University, Ulm, Germany

We investigate an ultraweak variational formulation for (parameterized) linear differential-algebraic equations with respect to 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 with respect to time in the spirit of the Reduced Basis Method. 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.

## OPEN ACCESS

### Edited by:

Benjamin Unger,  
University of Stuttgart, Germany

### Reviewed by:

Robert Altmann,  
University of Augsburg, Germany  
Martin Hess,  
International School for Advanced  
Studies (SISSA), Italy

### \*Correspondence:

Karsten Urban  
karsten.urban@uni-ulm.de

### Specialty section:

This article was submitted to  
Dynamical Systems,  
a section of the journal  
Frontiers in Applied Mathematics and  
Statistics

Received: 01 April 2022

Accepted: 31 May 2022

Published: 14 July 2022

### Citation:

Beurer E, Feuerle M, Reich N and  
Urban K (2022) An Ultraweak  
Variational Method for Parameterized  
Linear Differential-Algebraic Equations.  
Front. Appl. Math. Stat. 8:910786.  
doi: 10.3389/fams.2022.910786

**Keywords:** differential-algebraic equations, parametric equations, ultraweak formulations, Petrov-Galerkin methods, model reduction

## 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, refer to e.g., [1–4], or [5, 6], which are part of 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 [7–12].

All methods described in those references address a reduction of the dimension of the system, whereas the temporal discretization is untouched. This article 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 a right tool to achieve this goal. In [13], we have used this framework for deriving an optimally stable variational formulation of linear time-invariant systems (LTIs). In this article, we extend the ultraweak framework to (parameterized) DAEs and show that this can be combined with system theoretic methods such as Balanced Truncation (BT, [11]) to derive a reduction in the system dimension and time discretization size.

### 1.1. Differential-Algebraic Equations

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  $f: I \rightarrow \mathbb{R}^n$  a given right-hand side. Then, we are interested in the

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

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

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 exist some  $\hat{x}_0 \in \mathbb{R}^n$  such that  $E\hat{x}_0 - Ax_0 = \lim_{t \rightarrow 0^+} f(t)$  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  $\text{ind}\{E, -A\} = : k \in \mathbb{N}$ , [14].<sup>1</sup>

### 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  $x_\mu : I \rightarrow \mathbb{R}^n$  such that

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

where  $A_\mu, f_\mu$ , 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  $\mu$ , refer to footnote 5. 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, refer to e.g., [16]. 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

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

If such a decomposition is not given, we may produce an affinely decomposed approximation by means of the *(Discrete) Empirical Interpolation Method [(D)EIM]*, [17, 18]; refer also to [9] for a system theoretic MOR for such pDAEs]. For well-posedness, we assume that the matrix pencil  $\{E, -A_\mu\}$  is regular with index  $\text{ind}\{E, -A_\mu\} = k_\mu$  for all  $\mu \in \mathcal{P}$ .

### 1.3. Reduction to Homogeneous Initial Conditions

Using some standard arguments, eq. (1.1) can be reduced to homogeneous initial conditions  $x_\mu(0) = 0$ . To this end, construct some smooth extension of the initial data  $\tilde{x}_\mu \in C^1(\bar{I})^n$ ,  $\tilde{x}_\mu(0) =$

<sup>1</sup>Each regular matrix pencil can be transformed into *Weierstrass-Kronecker canonical form*  $P(\lambda E - A)Q = \text{diag}(\lambda I - W, \lambda N - Id)$  with regular matrices  $P, Q \in \mathbb{C}^{n \times n}$ , [15]. The index of a regular matrix pencil  $\{E, -A\}$  is then defined by  $\text{ind}\{E, -A\} := \text{ind}\{N\} := \min\{k \in \mathbb{N} : N^k = 0\}$ .

$x_{0,\mu}$ . Then, let  $\hat{x}_\mu : I \rightarrow \mathbb{R}^n$  solve eq. (1.1) with  $f_\mu$  replaced by  $\hat{f}_\mu := f_\mu - E\dot{\tilde{x}}_\mu + A_\mu \tilde{x}_\mu$  and homogeneous initial condition  $\hat{x}_\mu(0) = 0$ . Then,  $x_\mu := \hat{x}_\mu + \tilde{x}_\mu$  solves the original problem eq. (1.1). If the pDAE and the extension  $\tilde{x}_\mu$  of the initial conditions possess an affine decomposition (for a decomposable  $\tilde{x}_\mu$  refer to Section 3.2.2), 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.

### 1.4. 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 the results of our numerical experiments and end with conclusion 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., [3, 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) with respect to the time variable in the spirit of the reduced basis method, refer to Section 4 below.

### 2.1. An 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 [19] in which well-posed (ultraweak) variational forms for transport problems have been introduced, refer also to [20–22]. We will then transfer this framework to pDAEs in Section 2.2.

Let  $\Omega \subset \mathbb{R}^d$  be some open and bounded domain. We consider a *classical*<sup>2</sup> linear operator  $B_{\mu;\circ}$  on  $\Omega$  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

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

Note that the definition of  $B_{\mu;\circ}$  also incorporates essential homogeneous boundary conditions (in the 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

<sup>2</sup>By *classical* we mean defined in a pointwise manner.

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  $x, y \in C_0^\infty(\Omega)$  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  $B_{\mu;\circ} : D(B_{\mu;\circ}) \rightarrow R(B_{\mu;\circ})$  and  $B_{\mu;\circ}^* : D(B_{\mu;\circ}^*) \rightarrow R(B_{\mu;\circ}^*)$ . The following assumptions<sup>3</sup> 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

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

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

$$\begin{aligned} Y_\mu &:= \text{clos}_{|||\cdot|||_\mu}(D(B_{\mu;\circ}^*)), \\ (v, w)_{Y_\mu} &:= (B_{\mu;\circ}^*v, B_{\mu;\circ}^*w)_{L_2(\Omega)}, \\ \|v\|_{Y_\mu}^2 &:= (v, v)_{Y_\mu} = |||v|||_\mu^2, \end{aligned}$$

is a Hilbert space. Defining the bilinear form

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

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

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

Well-posedness including optimal stability is now ensured:

**Lemma 2.1.** *Problem eq. (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 respectively dual inf-sup constants read

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

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

*Proof:* Refer to [19, Proposition 2.1].

<sup>3</sup>The framework in [19] is slightly more general.

<sup>4</sup> $Y_\mu'$  denotes the dual space of  $Y_\mu$  with respect to the pivot space  $L_2(\Omega)$ .

## 2.2. An Ultraweak Formulation of pPDAEs

We are now going to apply the framework of Section 2.1 to the classical form (1.1) of the pDAE. Again, without loss of generality we restrict ourselves to homogeneous initial conditions  $x_\mu(0) = 0$ , as stated in Section 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

$$\begin{aligned} B_{\mu;\circ} &:= E \frac{d}{dt} - A_\mu, \\ D(B_{\mu;\circ}) &:= \{x \in C^1(I)^n \cap C(\bar{I})^n : x(0) = 0\}. \end{aligned}$$

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

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

which shows that

$$\begin{aligned} B_{\mu;\circ}^* &:= -E^T \frac{d}{dt} - A_\mu^T, \\ D(B_{\mu;\circ}^*) &\equiv C_E^1(I)^n := \{y \in C^1(I)^n \cap C(\bar{I})^n : y(T) \in \ker(E^T)\}. \end{aligned} \tag{2.3}$$

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 due to  $x(0) = 0$  and  $y(T) \in \ker(E^T)$ . Moreover,  $R(B_{\mu;\circ}) = C^{k_\mu-1}(I)^n$  and  $R(B_{\mu;\circ}^*) = C(I)^n$ .

**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.*

*Proof:* By the definition of  $H_0^1(I)^n$  and [23, 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,  $C_0^\infty(I)^n \subseteq D(B_{\mu;\circ}), D(B_{\mu;\circ}^*), R(B_{\mu;\circ}^*) \subset L_2(I)^n$  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

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

$$d_\mu(T) = y_\mu(T) - z_\mu(T) \in \ker(E^T).$$

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, [2, 3], i.e.,

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

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

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

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

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

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

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

Thus,  $u_\mu(T) = 0$  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, refer to e.g., [1]. 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 eq. (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

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

refer to (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  $\|y_\ell - y\|_{H^1(I)^n} \rightarrow 0$  as  $\ell \rightarrow \infty$ . Since  $\mathcal{P}$  is compact, we have that

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

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

The latter result must be properly interpreted. It says that  $Y_\mu$  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*

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

*If (1.1) admits a classical solution, then it coincides with  $x_\mu$ . 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  $x_\mu$  satisfying (2.6) is a weak solution to (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.,  $\lim_{\ell \rightarrow \infty} |||y - \tilde{y}_\ell|||_\mu = 0$ . Then, by the generalized Cauchy-Schwarz inequality

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

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 the context of (1.1) as a function of time, i.e.,  $g_\mu : I \rightarrow \mathbb{R}^n$ . Then, we simply define  $f_\mu \in Y'$  by

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

### 3. PETROV-GALERKIN DISCRETIZATION

The next step toward 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

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

which leads to solving a linear system of equations  $B_\mu^\mathcal{N} x_\mu^\mathcal{N} = f_\mu^\mathcal{N}$  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  $\|B_\mu^\mathcal{N} x_\mu^\mathcal{N} - f_\mu^\mathcal{N}\|^2 \rightarrow \min$ .

(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 eq. (3.1) is well-posed and optimally conditioned, [20], i.e.,

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

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

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

(c) The Xu-Zikatanov lemma [24] ensures that the Petrov-Galerkin error is comparable with the error of the best approximation, namely

$$\|x_\mu - x_\mu^\mathcal{N}\|_{L_2} \leq \frac{\gamma_\mu}{\beta_\mu^\mathcal{N}} \inf_{v^\mathcal{N} \in X_\mu^\mathcal{N}} \|x_\mu - v^\mathcal{N}\|_{L_2}, \quad (3.2)$$

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

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

Then, it is a standard estimate that

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

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

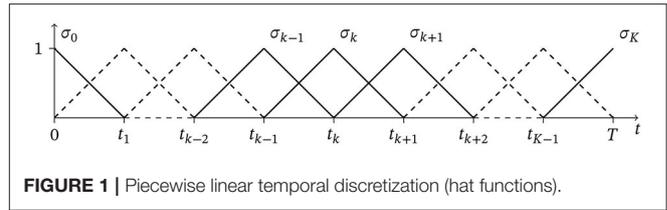


FIGURE 1 | Piecewise linear temporal discretization (hat functions).

#### 3.2. pPDAE 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  $t_k := k\Delta t, k = 0, \dots, K$  in  $I$ . Denote by  $\sigma_k, k = 0, \dots, K$  piecewise linear splines corresponding to the nodes  $t_{k-1}, t_k$  and  $t_{k+1}$ , refer to **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 [20], 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<sup>5</sup> and form a matrix  $V := (v_1, \dots, v_d) \in \mathbb{R}^{n \times d}$  by arranging the vectors as columns of  $V$ . Then, we construct  $Y^\mathcal{N} \subset Y = H_E^1(I)^n$  independent of the parameter and choose the trial space as  $X_\mu^\mathcal{N} := B_\mu^* Y^\mathcal{N}$ , which will then guarantee that  $\beta_\mu^\mathcal{N} = 1$ . We suggest a piecewise linear discretization by

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

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

$$X_\mu^\mathcal{N} := B_\mu^* Y^\mathcal{N} = \text{span}\{-E^T e_i \sigma_k - A_\mu^T e_i \sigma_k : k = 0, \dots, K - 1, i = 1, \dots, n\} \\ \oplus \text{span}\{-A_\mu^T v_i \sigma_K : i = 1, \dots, d\} \subset X = L_2(I)^n,$$

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

$$\|x_\mu - x_\mu^\mathcal{N}\|_{L_2} = \inf_{v^\mathcal{N} \in X_\mu^\mathcal{N}} \|x_\mu - v^\mathcal{N}\|_{L_2} = \| \|r(x_\mu^\mathcal{N})\| \|'_\mu = \Delta_\mu^\mathcal{N}. \quad (3.4)$$

##### 3.2.1. The Linear System

To construct the discrete linear system  $B_\mu^\mathcal{N} x_\mu^\mathcal{N} = f_\mu^\mathcal{N}$  we need bases  $\{\xi_1(\mu), \dots, \xi_\mathcal{N}(\mu)\}$  of  $X_\mu^\mathcal{N}$  and  $\{\psi_1, \dots, \psi_\mathcal{N}\}$  of  $Y^\mathcal{N}$ . The stiffness matrix  $B_\mu^\mathcal{N} \in \mathbb{R}^{\mathcal{N} \times \mathcal{N}}$  can be computed by  $[B_\mu^\mathcal{N}]_{j,i} := b_\mu(\xi_i(\mu), \psi_j) = (\xi_i(\mu), B_\mu^* \psi_j)_{L_2}$ . We recall that  $X_\mu^\mathcal{N} =$

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

$B_\mu^* Y^\mathcal{N}$ , which implies that  $\xi_i(\mu) = B_\mu^* \psi_i$ , so that  $[B_\mu^\mathcal{N}]_{j,i} = (B_\mu^* \psi_i, B_\mu^* \psi_j)_{L_2}$  and  $B_\mu^\mathcal{N}$  is in fact symmetric positive definite.

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

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

$$[K_{\Delta t}]_{k,\ell} := (\dot{\sigma}_k, \dot{\sigma}_\ell)_{L_2(I)},$$

$$[L_{\Delta t}]_{k,\ell} := (\sigma_k, \sigma_\ell)_{L_2(I)},$$

$$[O_{\Delta t}]_{k,\ell} := (\dot{\sigma}_k, \sigma_\ell)_{L_2(I)},$$

and subdivide the matrices  $\Pi_{\Delta t} \in \mathbb{R}^{(K+1) \times (K+1)}$  for  $\Pi_{\Delta t} \in \{K_{\Delta t}, L_{\Delta t}, O_{\Delta t}\}$  according to

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

Then, the stiffness matrix also has a block structure

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

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

$$\begin{aligned} B_\mu^{1,1} &= K_{\Delta t}^{1,1} \otimes EE^T + O_{\Delta t}^{1,1} \otimes EA_\mu^T + (O_{\Delta t}^{1,1})^T \otimes A_\mu E^T + L_{\Delta t}^{1,1} \\ &\quad \otimes A_\mu A_\mu^T \in \mathbb{R}^{nK \times nK}, \\ B_\mu^{1,2} &= O_{\Delta t}^{1,2} \otimes EA_\mu^T V + L_{\Delta t}^{1,2} \otimes A_\mu A_\mu^T V \in \mathbb{R}^{nK \times d}, \\ B_\mu^{2,1} &= (O_{\Delta t}^{2,1})^T \otimes V^T A_\mu E^T + L_{\Delta t}^{2,1} \otimes V^T A_\mu A_\mu^T \in \mathbb{R}^{d \times nK}, \\ B_\mu^{2,2} &= L_{\Delta t}^{2,2} \otimes V^T A_\mu A_\mu^T V \in \mathbb{R}^{d \times d}. \end{aligned}$$

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

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

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  $f_\mu$  and the test functions) and we choose the described one just for simplicity.

Due to the Kronecker product structure, the dimension of the system grows rapidly even for moderate  $n$  and  $K$ . The

efficient numerical solution thus requires a solver that takes the specific Kronecker product structure into account in order to accommodate the large system dimension. For similar Kronecker product systems arising from space-time (ultraweak) variational formulations of heat, transport, and wave equations, such specific efficient solvers have been introduced in [21, 25]. However, the solvers in these article are adapted to pPDEs and cannot be used for such kinds of pDAEs we are considering here. Hence, we will devote the development of efficient solvers for ultraweak formulations of pDAEs to future research and refer to Section 6.

### 3.2.2. Special Case: Fully Linear DAEs

We are going to specify the above general setting to the special case of DAEs with parameter-independent  $A \equiv A_\mu$  and linear right-hand side which we will call in the following *fully linear DAEs*, i.e.,

$$E\dot{x}_\mu(t) - Ax_\mu(t) = Du(t) + g_{\mu_2}(t), \quad t \in I, \quad x_\mu(0) = 0, \tag{3.5}$$

in which the right-hand side depends on the composite parameter  $\mu := (u, \mu_2)$  and is given in terms of a matrix  $D \in \mathbb{R}^{n \times m}$ , a control  $u: \bar{I} \rightarrow \mathbb{R}^m$ ,  $m$  denoting some input dimension and a function  $g_{\mu_2}: \bar{I} \rightarrow \mathbb{R}^n$ , which arises from the reduction to homogeneous initial conditions, refer to Section 1.3. The initial condition is assumed to be parameterized through  $g_{\mu_2}$  by  $\mu_2 \in \mathcal{P}_2 \subset \mathbb{R}^{p_2}$ ,  $p_2 \in \mathbb{N}$ . In view of (1.2) and Section 1.3, we get

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

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

We view the control and the initial condition (via  $g_{\mu_2}$ ) as parameters, i.e.,  $f_\mu(t) = D\mu_1(t) + g_{\mu_2}(t)$ ,  $\mu = (\mu_1, \mu_2)$ , 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 a parameter, i.e.,

$$\mu_1 := (u(t_0), \dots, u(t_K))^T \in \mathcal{P}_1 = \mathbb{R}^{p_1}, \quad p_1 = m(K+1),$$

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

$$f_\mu^\mathcal{N} = F_{\Delta t}^T (D \otimes Id_{K+1}) \mu_1 + \sum_{q=1}^{Q_x} \vartheta_q^x(\mu_2) F_{\Delta t}^T z_q,$$

so that the parameter dimension is  $p = p_1 + p_2 = m(K+1) + p_2$ , which might be large. The right-hand side  $f_\mu^\mathcal{N}$  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, we nevertheless have  $Q_f \ll \mathcal{N}$  if  $m \ll n$ .

Moreover, in the fully 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 with respect to 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, [10, 11], with the exception [9]. 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 that has originally been constructed for parameterized partial differential equations (pPDEs), refer to e.g., [7, 16, 26]. In an offline training phase, a reduced basis of size  $N \ll \mathcal{N}$  is constructed (typically in a greedy manner, refer to **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  $x_\mu$ , 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  $\mu$ , the reduced approximation  $x_N(\mu)$ <sup>6</sup> is then computed by solving a reduced system of dimension  $N$ . As a result of 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 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  $\mu$ . 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 with respect to  $K$  (both for parameterized and non-parameterized versions), so we even might assume that such Model Order Reduction (MOR) techniques have already been applied in a preprocessing step. We mention [9] a system theoretic MOR for parameter-dependent DAEs. Here, we are going to consider the reduction with respect to time using the RBM based upon a variational formulation with respect to 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

<sup>6</sup>For all quantities of the reduced system, we write the parameter  $\mu$  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.

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  $A_\mu$  can be treated similar to the RBM for ultraweak formulations of pPDEs as described e.g., in [20–22]. We further 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 with respect to the right-hand side.

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

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

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 Section 3.2 and obtain a reduced space of dimension  $N$  by setting

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

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  $\mu$ , 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

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

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

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

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

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

Its computation can be done in an online efficient manner in  $\mathcal{O}(N)$  operations by determining Riesz representations in the offline phase, refer to [7, 16, 26]. 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  $N_{\text{max}} \in \mathbb{N}$

- 1: choose  $\mu^{(1)} \in \mathcal{P}_{\text{train}}$ , compute snapshot  $\zeta_1 := x_{\mu^{(1)}}^{\mathcal{N}}$  and optimal test function  $\eta_1$  with  $\zeta_1 = B^* \eta_1$
- 2: **Initialize**  $S_1 \leftarrow \{\mu^{(1)}\}$ ,  $X_1 := \text{span}\{\zeta_1\}$ ,  $Y_1 := \text{span}\{\eta_1\}$ ,  $N := 1$
- 3: **while**  $N < N_{\text{max}}$  **do**
- 4:   **if**  $\max_{\mu \in \mathcal{P}_{\text{train}}} \Delta_N(\mu) \leq \varepsilon$  **then return**
- 5:    $\mu^{(N+1)} \leftarrow \arg \max_{\mu \in \mathcal{P}_{\text{train}}} \Delta_N(\mu)$
- 6:   compute snapshot  $\zeta_{N+1} := x_{\mu^{(N+1)}}^{\mathcal{N}}$  and optimal test function  $\eta_{N+1}$
- 7:    $S_{N+1} \leftarrow S_N \cup \{\mu^{(N+1)}\}$ ,  $X_{N+1} := X_N \oplus \text{span}\{\zeta_{N+1}\}$ ,  $Y_{N+1} := Y_N \oplus \text{span}\{\eta_{N+1}\}$
- 8:    $N \leftarrow N + 1$
- 9: **end while**

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

---

## 5. NUMERICAL EXPERIMENTS

In this section, we report on the 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, refer also to our remarks in Section 6 below. The code for producing the subsequent results is available via [https://github.com/mfeuerle/Ultraweak\\_PDAE](https://github.com/mfeuerle/Ultraweak_PDAE).

### 5.1. Serial RLC Circuit

We start with a standard problem which (in some cases) admits a closed formula for the analytical solution. This allows us to monitor the exact error and 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  $f_{V_S}: \bar{I} \rightarrow \mathbb{R}$ . Kirchhoff's circuit and further laws from electrical engineering yield a DAE with the data.

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

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 the resistor  $x_{V_R}$ .

#### 5.1.1. 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*, [27]) and by our ultraweak formulation from Section 3.2. We choose two specific examples for  $f_{V_S}$ , namely a smooth and a discontinuous one,

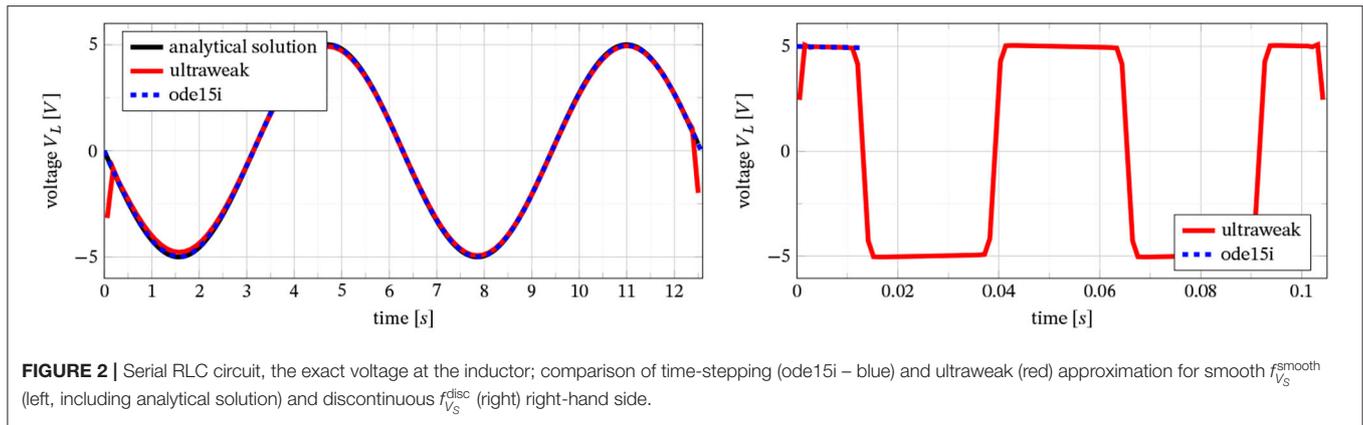
$$f_{V_S}^{\text{smooth}}(t) := \sin\left(\frac{4\pi}{T}t\right), \quad f_{V_S}^{\text{disc}}(t) := \text{sign}\left(\cos\left(\frac{4\pi}{T}t\right)\right).$$

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 pointwise comparisons are not necessarily meaningful.

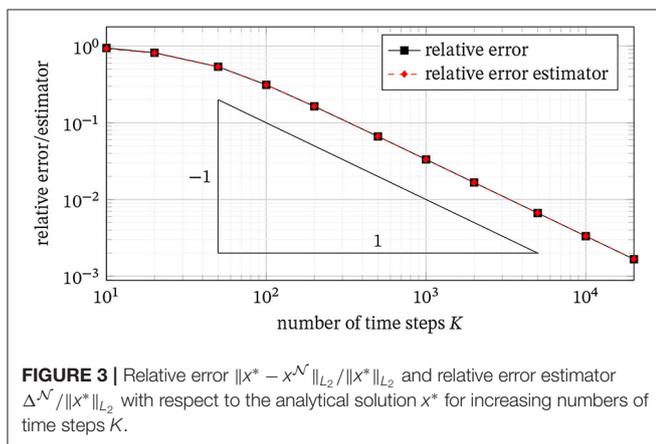
In the discontinuous case, the existence of a classical solution cannot be guaranteed by the above arguments. In particular, there is no closed solution formula. As we see in the right graph in **Figure 2**, *ode15i* stops at the first jump. This is to be expected, since  $f_{V_S}^{\text{disc}} \notin C^0(\bar{I})$ , 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.

#### 5.1.2. Convergence Rate

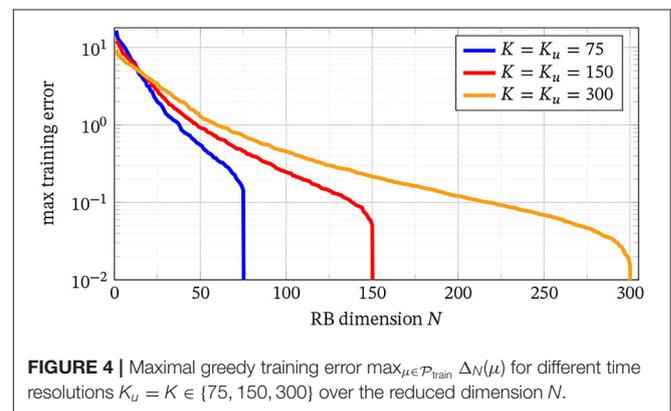
Next, we investigate the rate of convergence for the ultraweak form. To that end, we use  $f_{V_S}^{\text{smooth}}$ , since the analytical solution  $x^*$  is known and we can thus compute the relative error  $\|x^* - x^{\mathcal{N}}\|_{L_2} / \|x^*\|_{L_2}$ . Using the lowest order discretization as mentioned above (namely piecewise linear test functions  $\psi_j$ , which yield discontinuous trial functions  $B^* \psi_j$ ), we can only hope for the first order (with respect to the number of time



**FIGURE 2** | Serial RLC circuit, the 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.



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



**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  $N$ .

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.

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

### 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, [28–31], 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, [31], where we choose  $n = 644$ . The arising homogeneous fully linear DAE with output function  $y: I \rightarrow \mathbb{R}$  takes the form (3.5),

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

$$y(t) = Cx(t), \quad (5.1b)$$

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

In order to combine system theoretic model reduction with the RBM from Section 4, we use the system theoretic model order reduction package [28]. In particular, we use Balanced Truncation (BT) from [30] during a preprocessing step to reduce the above system of dimension  $n$  to a system

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

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

with  $\hat{E}, \hat{A} \in \mathbb{R}^{\hat{n} \times \hat{n}}$ ,  $\hat{D} \in \mathbb{R}^{\hat{n} \times 1}$ ,  $\hat{C} \in \mathbb{R}^{1 \times \hat{n}}$  as well as  $\hat{x}, \hat{g}: I \rightarrow \mathbb{R}^{\hat{n}}$  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 an LTI system, 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 [13].

*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  $\Delta_N^y$ , it is fairly standard in the theory of RBM to use a primal-dual approach with a second (dual) reduced basis, e.g., [16, 26]. For simplicity of exposition, we leave this to future study and compute the output from the state by  $\hat{C}\hat{x}(t)$ , respectively  $Cx(t)$ .

### 5.2.1. 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 Section 3.2 for different step sizes  $T/K_u$ , 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.

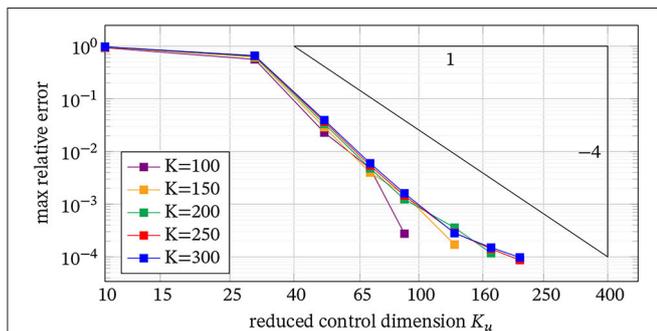
In order to answer this question, we apply **Algorithm 1** to the time-dependent Stokes problem (5.1a) (without BT) setting  $\varepsilon = 0$ ,  $N_{\max} = Q_f$  from (1.2) (i.e.,  $Q_f = p + 1$  for the fully linear system with parameter-independent initial value) and  $\mathcal{P}_{\text{train}}$  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 with respect to 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.

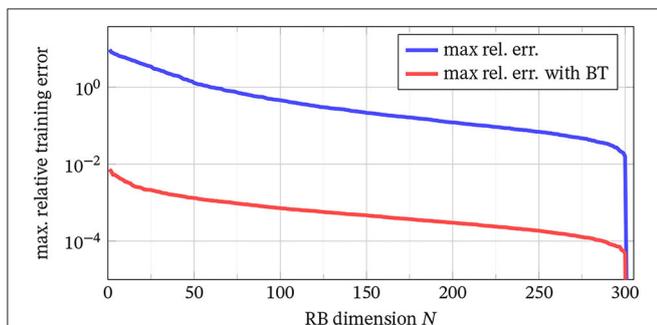
Next, we note that for  $A \equiv A_\mu$  as (5.1a), the reduced model is always *exact* for  $N \geq Q_f$ , which explains the drop off of the curves in **Figure 4**. For fully linear DAEs, a reduced model with  $N \geq Q_f = K_u + 2$  is always exact. Hence, if  $m \ll n$  (here  $m = 1 \ll 644 = n$ ), 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 an RBM seems necessary.

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 whether 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, [7, 16, 22], which is a property only of the problem at hand. In other words, if a pDAE can be reduced with respect to 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 with respect to 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  $K_u < K$ .



**FIGURE 6** | Maximal RBM relative error decay over the reduced dimension  $N$  for the full system eq. (5.1a) (blue) and the reduced system eq. (5.2a) with  $K = 300$  (red).

### 5.2.2. 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 with respect to time can be combined. To this end, we fix the temporal resolution (i.e., the number of time steps, here  $K = K_u = 300$ ) and determine the RBM error using **Algorithm 1** for the full and the BT-reduced system. We use [28] to compute the BT from [30] and obtain an LTI system of dimension  $\hat{n} = 5$ .

The results are shown in **Figure 6**, where we again show the maximal training error. As can be seen, the error for the BT-reduced system is smaller than the original one<sup>7</sup>, which in fact indicates that we can combine both methods. We get similar results for other choices of  $K$ . This shows that there is as much *reduction potential* in the reduced system eq. (5.2a) as in the original system eq. (5.1a). In other words, a combination of BT and RBM shows significant compression potential.

However, applying BT first implies that the error estimator measures the difference with respect to the BT-reduced model and not with respect to the original problem, which means that we somehow lose part of the rigor implied by the fact that our error estimator coincides with the error. If one wants to preserve the rigor, it seems that a full pPDE model is required, which is possible in cases where the pDAE originates from a (known) discretization of a pPDE.

## 6. CONCLUSION AND OUTLOOK

In this article, 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.

<sup>7</sup>This remains true even after additionally normalizing the training error by the dimensions of the DAE ( $n$  and  $\hat{n}$ , respectively) or the dimension of the resulting linear system ( $Kn + d$  and  $K\hat{n}$ , respectively).

## REFERENCES

- Brenan KE, Campbell SL, Petzold LR. *Numerical Solution of Initial-Value Problems in Differential-Algebraic Equations*. Philadelphia, PA: SIAM (1995). doi: 10.1137/1.9781611971224
- Hairer E, Wanner G. *Solving Ordinary Differential Equations II*. Vol. 14. Berlin: Springer (2010).
- Kunkel P, Mehrmann V. *Differential-Algebraic Equations*. Zürich: European Mathematical Society (EMS) (2006).
- Trenn S. In: Ilchmann A, Reis T. *Solution Concepts for Linear DAEs: A Survey*. Berlin: Springer (2013). doi: 10.1007/978-3-642-34928-7\_4
- Grundel S, Reis T, Schöps S. *Progress in Differential-Algebraic Equations II*. Cham: Springer (2020).
- Schöps S, Bartel A, Günther M, ter Maten EJ, Müller PC. *Progress in Differential-Algebraic Equations*. Berlin: Springer (2014).
- Benner P, Cohen A, Ohlberger M, Willcox K. *Model Reduction and Approximation: Theory and Algorithms*. Philadelphia, PA: SIAM (2017). doi: 10.1137/1.9781611974829
- Benner P, Stykel T. Model order reduction for differential-algebraic equations: A survey. In: Ilchmann A, Reis T, editors. *Surveys in Differential-Algebraic Equations IV*. Cham: Springer (2017). p. 107–60. doi: 10.1007/978-3-319-46618-7\_3
- Chellappa S, Feng L, Benner P. Adaptive basis construction and improved error estimation for parametric nonlinear dynamical systems. *Int J Num Methods Eng*. (2020) 121:5320–49. doi: 10.1002/nme.6462
- Mehrmann V, Stykel T. Balanced truncation model reduction for large-scale systems in descriptor form. In: Benner P, Sorensen DC, Mehrmann V, editors. *Dimension Reduction of Large-Scale Systems*. Berlin: Springer (2005). p. 83–115. doi: 10.1007/3-540-27909-1\_3
- Stykel T. Balanced truncation model reduction for descriptor systems. *PAMM*. (2003) 3:5–8. doi: 10.1002/pamm.200310302
- Stykel T. Gramian-based model reduction for descriptor systems. *Math Control Signals Syst*. (2004) 16:297–319. doi: 10.1007/s00498-004-0141-4
- Feuerle M, Urban K. A variational formulation for LTI-systems and model reduction. In: Vermolen FJ, Vuik C, editors. *Numerical Mathematics and Advanced Applications ENUMATH 2019*. Cham: Springer (2021). p. 1059–67. doi: 10.1007/978-3-030-55874-1\_105
- Gear CW, Petzold LR. Differential/algebraic systems and matrix pencils. In: Kragström B, Ruhe A, editors. *Matrix Pencils*. Berlin: Springer (1983). p. 75–89. doi: 10.1007/BFb0062095
- Gantmacher FR. *Theory of Matrices II*. Providence, RI: Chelsea Publication (1959).
- Hesthaven JS, Rozza G, Stamm B. *Certified Reduced Basis Methods for Parametrized Partial Differential Equations*. Cham: Springer (2016). doi: 10.1007/978-3-319-22470-1
- Barrault M, Maday Y, Nguyen NC, Patera AT. An empirical interpolation method: application to efficient reduced-basis discretization of partial differential equations. *C R Acad Sci*. (2004) 339:667–72. doi: 10.1016/j.crma.2004.08.006
- Chaturantabut S, Sorensen D. Nonlinear model reduction via discrete empirical interpolation. *SIAM J Sci Comput*. (2010) 32:2737–64. doi: 10.1137/090766498

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 with respect to 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  $A_\mu$ , 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 [21] 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.

## DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

## AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

## ACKNOWLEDGMENTS

We acknowledge support by the state of Baden-Württemberg through bwHPC.

19. Dahmen W, Huang C, Schwab C, Welper G. Adaptive Petrov-Galerkin methods for first order transport equations. *SIAM J Num Anal.* (2012) 50:2420–45. doi: 10.1137/110823158
20. Brunken J, Smetana K, Urban K. (Parametrized) First order transport equations: realization of optimally stable Petrov-Galerkin methods. *SIAM J Sci Comput.* (2019) 41:A592–621. doi: 10.1137/18M1176269
21. Henning J, Palitta D, Simoncini V, Urban K. An ultraweak space-time variational formulation for the wave equation: Analysis and efficient numerical solution. *ESAIM: Math Modell Numer Anal.* (2022) 56:1173–98. Available online at: <https://www.esaim-m2an.org/articles/m2an/pdf/2022/04/m2an210138.pdf>
22. Urban K. *The Reduced Basis Method in Space and Time: Challenges, Limits and Perspectives.* Ulm: Ulm University (2022).
23. Arendt W, Urban K. *Partial Differential Equations: An analytic and Numerical Approach.* Transl. by J. B. Kennedy. New York, NY: Springer (2022).
24. Xu J, Zikatanov L. Some observations on Babuška and Brezzi theories. *Num Math.* (2003) 94:195–202. doi: 10.1007/s002110100308
25. Henning J, Palitta D, Simoncini V, Urban K. Matrix oriented reduction of space-time Petrov-Galerkin variational problems. In: Vermolen FJ, Vuik C, editors. *Numerical Mathematics and Advanced Applications ENUMATH 2019.* Cham: Springer (2019). p. 1049–57. doi: 10.1007/978-3-030-55874-1\_104
26. Quarteroni A, Manzoni A, Negri F. *Reduced Basis Methods for Partial Differential Equations: An Introduction.* vol. 92. Springer (2015). doi: 10.1007/978-3-319-15431-2
27. The MathWorks, Inc. *MATLAB Version 9.11.0.* Natick, MA (2021).
28. Saak J, Köhler M, Benner P. *MM.ESS-2.2-The Matrix Equations Sparse Solvers Library.* (2022). Available online at: <https://www.mpi-magdeburg.mpg.de/projects/mess>
29. Schmidt M. *Systematic Discretization of Input/Output Maps and Other Contributions to the Control of Distributed Parameter Systems.* Berlin: TU Berlin (2007).
30. Stykel T. Balanced truncation model reduction for semidiscretized Stokes equation. *Linear Algebra Appl.* (2006) 415:262–289. doi: 10.1016/j.laa.2004.01.015
31. The MORwiki Community. *Stokes Equation.* Magdeburg: MORwiki-Model Order (2018).

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Beurer, Feuerle, Reich and Urban. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.