



An adaptive certified space-time reduced basis method for nonsmooth parabolic partial differential equations

Marco Bernreuther¹ · Stefan Volkwein¹

Received: 2 January 2023 / Accepted: 4 April 2024
© The Author(s) 2024

Abstract

In this paper, a nonsmooth semilinear parabolic partial differential equation (PDE) is considered. For a reduced basis (RB) approach, a space-time formulation is used to develop a certified a-posteriori error estimator. This error estimator is adopted to the presence of the discrete empirical interpolation method (DEIM) as approximation technique for the nonsmoothness. The separability of the estimated error into an RB and a DEIM part then guides the development of an adaptive RB-DEIM algorithm, combining both offline phases into one. Numerical experiments show the capabilities of this novel approach in comparison with classical RB and RB-DEIM approaches.

Keywords Nonsmooth parabolic equations · Space-time discretization · Reduced basis · Discrete empirical interpolation · A-posteriori error estimation · Semismooth Newton

Mathematics Subject Classification (2010) 35K55 · 65M15 · 65M50 · 65N30

1 Introduction

The usual approach for the numerical solution of parabolic partial differential equations (PDEs) is done by time-stepping schemes based upon variational semi-discretizations. After a variational formulation in space and a discretization by, e.g., finite elements, one derives an evolution problem in time. Then, a spatial problem needs to be solved at each discrete time instance.

Communicated by: Gianluigi Rozza

✉ Marco Bernreuther
marco.bernreuther@uni-konstanz.de
Stefan Volkwein
stefan.volkwein@uni-konstanz.de

¹ Department of Mathematics and Statistics, University of Konstanz, 78457 Konstanz, Germany

Alternatively, in a space-time scheme one approximates the PDE by a simultaneous discretization of the spatial *and* temporal domain. This leads to a single variational problem with test functions depending on the spatial and the temporal variables \mathbf{x} and t , respectively.

In this paper, we consider a parameter dependent, nonsmooth parabolic PDE: For a parameter $\mu \in \mathcal{P}_{\text{ad}}$ consider for almost all (f.a.a.) $t \in (0, T]$:

$$\begin{aligned} \dot{y}(t; \mu) - c(\mu)\Delta y(t; \mu) + a(\mu) \max\{0, y(t; \mu)\} &= f(t; \mu) \text{ in } V', \\ y(0; \mu) &= 0 \quad \text{in } H, \end{aligned} \quad (\mu\text{IVP})$$

where $\dot{y} = \frac{\partial y}{\partial t}$ denotes the weak temporal derivative of y . Notice that the nonsmoothness in (μIVP) is only Lipschitz continuous and not Fréchet-differentiable. Problem (μIVP) can be seen as a model problem for a broader class of nonsmooth parabolic PDEs, as shown in the first authors doctoral thesis [1], where the assumptions on the nonsmoothness are chosen appropriately. Note that in this general setting, the derivation of stability estimates, see Proposition 1 and Corollary 3 can be problematic. Let us mention that some examples of problems related to the nonsmooth PDE arise, for instance, in mechanics, plasma physics and the context of certain combustion processes, cf. [2, 3].

Let us mention some of the related work. Space-time methods have been considered by many authors. We refer, e.g., to the work [4–11] for (smooth) parabolic problems but there is no error analysis done for nonsmooth PDEs. In the context of reduced basis (RB) approaches space-time methods are discussed, e.g., in [12–15] and in particular for optimal control problems in [16–18]. A-posteriori error estimates are derived and efficient tensor-based solution methods are proposed. However, the authors do not study empirical interpolation methods and adaptive basis update strategies in their work. Empirical interpolation techniques are necessary to handle the nonsmooth term in (μIVP) efficiently by the RB method, cf. [19–21]. Theoretical results for (μIVP) can be found in [22, 23], where optimal control problems for more general parabolic nonsmooth problems are considered. In the context of adaptive RB methods we refer to [24], for instance. Let us also mention that the present paper extends results of the elliptic case (cf. [25]) to the parabolic one, which further justifies the consideration of the max-type nonsmoothness as a starting point.

The new contributions of the present paper are as follows: (i) We derive a-posteriori error estimates for space-time approximations of a *nonsmooth* parabolic PDEs. (ii) We incorporate the *discrete empirical interpolation method* (DEIM) for the nonsmooth term in our error analysis. (iii) A *certified adaptive algorithm* for the RB-DEIM approximation is developed, which combines the typical two offline phases for the computation of an RB and of a DEIM basis into one.

This paper is organized as follows: In Section 2, space-time formulations are introduced for the continuous, finite element (FE) and RB formulations of (μIVP) . Section 3 covers the RB method, including basis generation and space-time a-posteriori error estimation. In Section 4, DEIM is introduced to efficiently evaluate and approximate

the nonsmoothness. The error estimator is adopted to the RB-DEIM setting and an adaptive algorithm for the simultaneous generation of the RB and the DEIM basis is presented. Numerical results illustrate the capabilities this novel approach in Section 5. Finally, we draw some conclusions in Section 6.

2 Space-time formulation for (μ IVP)

First we establish a space-time formulation for the continuous setting and existence of a unique solution is proved. Then, an FE space-time discretization is introduced. Further, it is shown that its space-time formulation corresponds to a Crank-Nicolson (CN) scheme, which can be uniquely solved by a semismooth Newton method. Finally the same is done for the RB space-time formulation. We also present stability estimates for all formulations, that will be necessary to provide convergence rates for some of the RB error estimators quantities.

2.1 Problem formulation

Let $\Omega \subset \mathbb{R}^d, d \in \{1, 2, 3\}$, be a bounded domain with Lipschitz-continuous boundary $\Gamma = \partial\Omega$. We write $\mathbf{x} = (x_1, \dots, x_d)$ for an element in Ω . For $T > 0$ we define $Q = (0, T) \times \Omega$ and $\Sigma = (0, T) \times \Gamma$. Let $H = L^2(\Omega)$ and $V = H_0^1(\Omega)$ be supplied by the inner products

$$\begin{aligned} \langle \varphi, \phi \rangle_H &= \int_{\Omega} \varphi(\mathbf{x})\phi(\mathbf{x}) \, d\mathbf{x} && \text{for } \varphi, \phi \in H, \\ \langle \varphi, \phi \rangle_V &= \int_{\Omega} \nabla\varphi(\mathbf{x}) \cdot \nabla\phi(\mathbf{x}) \, d\mathbf{x} && \text{for } \varphi, \phi \in V, \end{aligned}$$

respectively, and their corresponding induced norms. Moreover, their dual spaces are denoted as H' and V' . Furthermore, we introduce the test space $\mathcal{Y} = L^2(0, T; V)$. We identify the dual \mathcal{Y}' with the space $L^2(0, T; V')$. For more details on Sobolev and Bochner spaces we refer the reader to [26], for instance.

Recall that $V \hookrightarrow H \simeq H' \hookrightarrow V'$ is a Gelfand triple and the function space $W(0, T) = \mathcal{Y} \cap H^1(0, T; V')$ a Hilbert space with induced norm

$$\|\varphi\|_{W(0,T)}^2 = \|\varphi\|_{\mathcal{Y}}^2 + \|\dot{\varphi}\|_{\mathcal{Y}'}^2 \quad \text{for } \varphi \in W(0, T).$$

Moreover, the solution space is $\mathcal{X} = \{\varphi \in W(0, T) \mid \varphi(0) = 0 \text{ in } H\}$ with norm

$$\|\varphi\|_{\mathcal{X}}^2 = \|\varphi\|_{W(0,T)}^2 + \|\varphi(T)\|_H^2 \quad \text{for } \varphi \in \mathcal{X},$$

which is well-defined due to $W(0, T) \hookrightarrow C([0, T]; H)$, cf. [27]. Finally, we set $\mathcal{H} = L^2(0, T; H)$ for brevity. For a function $\varphi \in \mathcal{H}$ we write $\varphi(t)$ for the function $\mathbf{x} \mapsto \varphi(t, \mathbf{x})$ for almost all (abbreviated “f.a.a.” in the following) $t \in [0, T]$. The precise assumptions on (μ IVP) are stated in the following assumption.

- Assumption 1** a) $\mathcal{P}_{\text{ad}} \subsetneq \mathbb{R}^p$, $p \in \mathbb{N} \setminus \{0\}$, is nonempty and compact,
 b) $c: \mathcal{P}_{\text{ad}} \rightarrow \mathbb{R}$ is Lipschitz-continuous, positive and uniformly bounded away from zero,
 c) $a: \mathcal{P}_{\text{ad}} \rightarrow \mathbb{R}$ is Lipschitz-continuous and nonnegative,
 d) $f: \mathcal{P}_{\text{ad}} \rightarrow C([0, T]; H)$ is Lipschitz-continuous.

Remark 1 Assumption 1-d) can be relaxed to $f: \mathcal{P}_{\text{ad}} \rightarrow \mathcal{H}$. However, we suppose that $t \mapsto f(t; \mu) \in H$ is continuous to simplify our presentation regarding the temporal discretization carried out later. \diamond

For $y \in \mathcal{X}$, $\phi \in \mathcal{Y}$ and $\mu \in \mathcal{P}_{\text{ad}}$ we define the operators

$$\begin{aligned} \mathcal{B}(y, \phi; \mu) &= \mathcal{B}_1(y, \phi) + c(\mu)\mathcal{B}_2(y, \phi), & \mathcal{B}_1(y, \phi) &= \langle \dot{y}, \phi \rangle_{\mathcal{Y}'} \mathcal{Y}, \\ \mathcal{A}(y, \phi; \mu) &= \mathcal{B}(y, \phi; \mu) + \mathcal{N}(y, \phi; \mu), & \mathcal{B}_2(y, \phi) &= \langle y, \phi \rangle_{\mathcal{Y}}, \\ \mathcal{N}(y, \phi; \mu) &= a(\mu) \langle \max\{0, y\}, \phi \rangle_{\mathcal{H}}, & \mathcal{F}(\phi; \mu) &= \langle f(\mu), \phi \rangle_{\mathcal{H}}. \end{aligned} \tag{1}$$

For $\mu \in \mathcal{P}_{\text{ad}}$ the function $y = y(\mu) \in \mathcal{X}$ is called a *weak solution* to (μ IVP) if

$$\mathcal{A}(y, \phi; \mu) = \mathcal{F}(\phi; \mu) \quad \text{for all } \phi \in \mathcal{Y}. \tag{2}$$

Existence and uniqueness of the solution to (2) follows e.g. from [28, Theorem 30.A].

2.2 FE space-time formulation

Analogously to the continuous setting we introduce a discretized space-time formulation. Therefore let $\mathcal{X}_\delta \subset \mathcal{X}$ and $\mathcal{Y}_\delta \subset \mathcal{Y}$ be finite dimensional subspaces. For $\mu \in \mathcal{P}_{\text{ad}}$ we call $y_\delta = y_\delta(\mu) \in \mathcal{X}_\delta$ a discretized weak solution to (μ IVP) if

$$\mathcal{A}(y_\delta, \phi_\delta; \mu) = \mathcal{F}(\phi_\delta; \mu) \quad \text{for all } \phi_\delta \in \mathcal{Y}_\delta. \tag{3}$$

In the remainder of this paper we will focus on the case

$$\mathcal{X}_\delta = S_{\Delta t} \otimes V_h, \quad \mathcal{Y}_\delta = Q_{\Delta t} \otimes V_h,$$

where \otimes denotes the tensor product and $Q_{\Delta t}$, $S_{\Delta t}$ are piecewise constant, respective piecewise linear finite elements in time and V_h are piecewise linear finite elements in space. As $\delta = (\Delta t, h)$ we summarize the temporal and spatial discretization parameters. The solutions for this particular choice of spaces will be also called FE solutions to the PDE. For the FE spaces we denote the standard bases (see [14]) by

$$Q_{\Delta t} = \text{span} \{ \tau_1, \dots, \tau_K \}, \quad S_{\Delta t} = \text{span} \{ \sigma_1, \dots, \sigma_K \}, \quad V_h = \text{span} \{ \zeta_1, \dots, \zeta_N \},$$

where $K \in \mathbb{N}$ and $N \in \mathbb{N}$ denote the sizes of the temporal and the spatial discretization, respectively. We set $I = [0, T)$ and introduce the intervals $I_k = [t_{k-1}, t_k)$ of length

Δt_k for the time instances $0 = t_0 < \dots < t_K = T$, i.e., $I = I_1 \cup \dots \cup I_K$. From

$$y_\delta = \sum_{k=1}^K \sum_{i=1}^N y_i^k(\sigma_k \otimes \zeta_i) \in \mathcal{X}_\delta, \quad \phi_\delta = \sum_{k=1}^K \sum_{i=1}^N \phi_i^k(\tau_k \otimes \zeta_i) \in \mathcal{Y}_\delta$$

we infer that

$$\begin{aligned} \mathcal{B}(y_\delta, \phi_\delta; \mu) &= \sum_{k,l=1}^K \sum_{i,j=1}^N y_j^l \phi_i^k \left(\langle \dot{\sigma}_l, \tau_k \rangle_{L^2(0,T)} \langle \zeta_j, \zeta_i \rangle_H \right. \\ &\quad \left. + c(\mu) \langle \sigma_l, \tau_k \rangle_{L^2(0,T)} \langle \zeta_j, \zeta_i \rangle_V \right) = y_\delta^\top \mathbf{B}(\mu) \phi \end{aligned}$$

with $y_\delta = [y_\delta^1 | \dots | y_\delta^K] \in \mathbb{R}^{N \times K}$, $y_\delta^k = (y_1^k, \dots, y_N^k)^\top$ for $1 \leq k \leq K$, $\phi = [\phi^1 | \dots | \phi^K] \in \mathbb{R}^{N \times K}$, $\phi^k = (\phi_1^k, \dots, \phi_N^k)^\top$ for $1 \leq k \leq K$ and $\mathbf{B}(\mu) = \mathbf{B}_1 + c(\mu)\mathbf{B}_2$. Here we have introduced spatio-temporal matrices $\mathbf{B}_1 = \mathbf{N}_h^{\text{time}} \otimes \mathbf{M}_h^{\text{space}}$ and $\mathbf{B}_2 = \mathbf{M}_h^{\text{time}} \otimes \mathbf{V}_h^{\text{space}}$ with

$$\begin{aligned} \mathbf{M}_h^{\text{space}} &= \left(\langle \zeta_j, \zeta_i \rangle_H \right)_{1 \leq i, j \leq N}, & \mathbf{V}_h^{\text{space}} &= \left(\langle \zeta_j, \zeta_i \rangle_V \right)_{1 \leq i, j \leq N}, \\ \mathbf{M}_h^{\text{time}} &= \left(\langle \dot{\sigma}_l, \tau_k \rangle_{L^2(0,T)} \right)_{1 \leq k, l \leq K}, & \mathbf{N}_h^{\text{time}} &= \left(\langle \sigma_l, \tau_k \rangle_{L^2(0,T)} \right)_{1 \leq k, l \leq K}. \end{aligned}$$

Let $\delta_{l,k}$ denote the Kronecker delta, then we obtain the explicit forms

$$\langle \dot{\sigma}_l, \tau_k \rangle_{L^2(0,T)} = \delta_{l,k} - \delta_{l+1,k}, \quad \langle \sigma_l, \tau_k \rangle_{L^2(0,T)} = \frac{1}{2} (\Delta t_l \delta_{l,k} + \Delta t_{l+1} \delta_{l+1,k}).$$

Furthermore we denote the lumped version of the spatial mass matrix $\mathbf{M}_h^{\text{space}}$ as

$$\tilde{\mathbf{M}}_h^{\text{space}} = \text{diag} \left(\frac{1}{3} [|\text{supp}(\zeta_j)|]_{1 \leq j \leq N} \right),$$

cf., e.g. [29, Chapter 15]. This leads to

$$\begin{aligned} \mathcal{B}(y_\delta, \tau_k \otimes \zeta_i; \mu) &= \left[\mathbf{M}_h^{\text{space}} \left(y_\delta^k - y_\delta^{k-1} \right) + \frac{c(\mu)\Delta t_k}{2} \mathbf{V}_h^{\text{space}} \left(y_\delta^k + y_\delta^{k-1} \right) \right]_i, \\ \mathcal{N}(y_\delta, \tau_k \otimes \zeta_i; \mu) &\approx \left[\frac{a(\mu)\Delta t_k}{2} \tilde{\mathbf{M}}_h^{\text{space}} \left(\max \{0, y_\delta^k\} + \max \{0, y_\delta^{k-1}\} \right) \right]_i, \\ \mathcal{F}(\tau_k \otimes \zeta_i; \mu) &\approx \frac{\Delta t_k}{2} \left(\mathbf{F}_i^k(\mu) + \mathbf{F}_i^{k-1}(\mu) \right) \end{aligned}$$

for $1 \leq k \leq K$ and $1 \leq i \leq N$, where we have used the trapezoidal quadrature rule for the approximation of the integrals in \mathcal{N} and \mathcal{F} and denote $\mathbf{F}(\mu) = [\mathbf{F}^0(\mu) | \dots | \mathbf{F}^K(\mu)] \in \mathbb{R}^{N \times K+1}$ with $\mathbf{F}^k(\mu) = (\mathbf{F}_1^k(\mu), \dots, \mathbf{F}_N^k(\mu))^\top$ and $\mathbf{F}_i^k(\mu) = \langle f(t_k; \mu), \zeta_i \rangle_{V',V}$, where we have utilized that Assumption 1-d) holds. Furthermore we set $y_\delta^0 = \mathbf{0} \in \mathbb{R}^N$ due

to the homogeneous initial condition. Now we can express (3) as a sequence of root finding problems for $k = 1, \dots, K$:

$$G_\delta^k(y_\delta^k; \mu) = 0 \quad \text{in } \mathbb{R}^N$$

with

$$G_\delta^k(y_\delta^k; \mu) = \frac{1}{\Delta t_k} M_h^{\text{space}}(y_\delta^k - y_\delta^{k-1}) + \frac{1}{2} [c(\mu) V_h^{\text{space}}(y_\delta^k + y_\delta^{k-1}) + a(\mu) \tilde{M}_h^{\text{space}}(\max\{0, y_\delta^k\} + \max\{0, y_\delta^{k-1}\}) - (F^k(\mu) + F^{k-1}(\mu))]$$

and initial condition $y_\delta^0 = 0$. This problem can be interpreted as a Crank-Nicolson (CN) scheme for a spatially discretized parabolic PDE. To solve this problem, a semismooth Newton method (see [30]), where the k -th iteration matrix is given by

$$H_\delta^k(y_\delta^k; \mu) = \frac{1}{\Delta t_k} M_h^{\text{space}} + \frac{1}{2} [c(\mu) V_h^{\text{space}} + a(\mu) \tilde{M}_h^{\text{space}} \Theta(y_\delta^k)] \in \mathbb{R}^{N \times N}$$

is applied. The function $\Theta: \mathbb{R}^N \rightarrow \mathbb{R}^{N \times N}$ maps a vector to the diagonal matrix that takes the Heaviside function with value 0 evaluated for each entry of the vector as its diagonal entries.

Remark 2 For every parameter $\mu \in \mathcal{P}_{\text{ad}}$ the problem $G_\delta^k(y_\delta^k; \mu) = 0$ for $k = 1, \dots, K$ with initial condition $y_\delta^0 = 0$ admits a unique sequence of roots y_δ^k . This is a consequence of the monotonicity of $G_\delta^k(\cdot; \mu)$, which follows since $M_h^{\text{space}}, V_h^{\text{space}}$ and $\tilde{M}_h^{\text{space}}$ are symmetric and positive definite (s.p.d.) matrices and the max-function is monotone. Furthermore for every y_δ^k the matrix $H_\delta^k(y_\delta^k; \mu)$ is s.p.d., which implies that every Newton iteration is uniquely solvable. \diamond

We finish our considerations with a stability estimate for the FE solution. This is based on the linear case shown in [4, Corollary 4.3].

Proposition 1 *Let Assumption 1 hold. Then, for every $\mu \in \mathcal{P}_{\text{ad}}$, the FE solution $y_\delta(\mu)$ to (3) satisfies the estimate*

$$\|\dot{y}_\delta(\mu)\|_{\mathcal{H}} \leq \|f(\mu)\|_{\mathcal{H}}. \tag{4}$$

Proof Let $y_\delta = y_\delta(\mu)$ be an FE solution to (3). Since $\dot{y}_\delta \in \mathcal{Y}_\delta$ holds, we can choose $\phi_\delta = \dot{y}_\delta$ as test function. This implies

$$\langle \dot{y}_\delta, \dot{y}_\delta \rangle_{\mathcal{H}} + c(\mu) \langle y_\delta, \dot{y}_\delta \rangle_{\mathcal{Y}} + a(\mu) \langle \max\{0, y_\delta\}, \dot{y}_\delta \rangle_{\mathcal{H}} = \langle f(\mu), \dot{y}_\delta \rangle_{\mathcal{H}}.$$

Note that

$$c(\mu)\langle y_\delta, \dot{y}_\delta \rangle_{\mathcal{Y}} = \frac{c(\mu)}{2} (\|y_\delta(T)\|_V^2 - \|y_\delta(0)\|_V^2) = \frac{c(\mu) \|y_\delta(T)\|_V^2}{2} \geq 0,$$

$$\langle f(\mu), \dot{y}_\delta \rangle_{\mathcal{H}} \leq \frac{1}{2} (\|f(\mu)\|_{\mathcal{H}}^2 + \|\dot{y}_\delta\|_{\mathcal{H}}^2).$$

Thus we obtain

$$\|\dot{y}_\delta\|_{\mathcal{H}}^2 + 2a(\mu) \langle \max\{0, y_\delta\}, \dot{y}_\delta \rangle_{\mathcal{H}} \leq \|f(\mu)\|_{\mathcal{H}}^2.$$

Now for almost all $t \in (0, T)$ we introduce the set $\Omega^+(t) = \{\mathbf{x} \in \Omega \mid y_\delta(t, \mathbf{x}) > 0 \text{ f.a.a. } t \in [0, T]\}$. Since $\max\{0, \cdot\}: \mathbb{R} \rightarrow \mathbb{R}$ is Lipschitz continuous, we infer from [26, Chapter 5.8.2.b] that $\max\{0, y_\delta(t)\}$ belongs to V f.a.a. $t \in (0, T)$ and thus

$$\begin{aligned} \int_I \int_{\Omega^+(t)} y_\delta \dot{y}_\delta \, d\mathbf{x} dt &= \int_I \int_{\Omega} \max\{0, y_\delta\} \dot{y}_\delta \, d\mathbf{x} dt \\ &= \int_{\Omega} \max\{0, y_\delta(T)\} y_\delta(T) \, d\mathbf{x} - \int_{\Omega} \max\{0, y_\delta(0)\} y_\delta(0) \, d\mathbf{x} \\ &\quad - \int_I \int_{\Omega} \max'\{0, y_\delta\} \dot{y}_\delta y_\delta \, d\mathbf{x} dt \\ &= \int_{\Omega} \max\{0, y_\delta(T)\} y_\delta(T) \, d\mathbf{x} - \int_I \int_{\Omega^+(t)} y_\delta \dot{y}_\delta \, d\mathbf{x} dt, \end{aligned}$$

since for the weak derivative $\max'\{0, y_\delta\} = 1$ almost everywhere (abbreviated ‘‘a.e.’’ in the following) on $\Omega^+(t)$ and $\max'\{0, y_\delta\} = 0$ a.e. on $\Omega \setminus \Omega^+(t)$ are satisfied; cf. [31] for a derivation based on generalized derivatives. This implies

$$\langle \max\{0, y_\delta\}, \dot{y}_\delta \rangle_{\mathcal{H}} = \frac{1}{2} \langle \max\{0, y_\delta(T)\}, y_\delta(T) \rangle_H \geq 0$$

and thus (4) holds true. □

For an arbitrarily given $\varphi_\delta \in \mathcal{X}_\delta$, which is piecewise linear in time, let us define the piecewise constant in time version by averaging

$$\bar{\varphi}_\delta(t, \mathbf{x}) = \sum_{k=1}^K \mathbb{1}_{I_k}(t) \otimes \bar{\varphi}_\delta^k(\mathbf{x}) \in \mathcal{Y}_\delta \tag{5}$$

with

$$\bar{\varphi}_\delta^k = \frac{1}{\Delta t_k} \int_{I_k} \varphi_\delta(t) \, dt = \frac{\varphi(t_{k-1}) + \varphi(t_k)}{2} \in V. \tag{6}$$

In the next lemma we summarize useful properties of $\dot{\varphi}_\delta \in \mathcal{H}$. Recall that $\dot{\varphi}_\delta \in \mathcal{H}$ for $\varphi_\delta \in \mathcal{X}_\delta$.

Lemma 2 For every $\varphi_\delta \in \mathcal{X}_\delta$ we have

$$\langle \dot{\varphi}_\delta, \bar{\varphi}_\delta \rangle_{\mathcal{H}} = \frac{1}{2} \|\varphi_\delta(T)\|_H^2, \tag{7}$$

and

$$\langle \varphi_\delta - \bar{\varphi}_\delta, \phi_\delta \rangle_{\mathcal{Y}} = \langle \varphi_\delta - \bar{\varphi}_\delta, \phi_\delta \rangle_{\mathcal{H}} = 0 \text{ for all } \phi_\delta \in \mathcal{Y}_\delta. \tag{8}$$

Proof Let $\varphi_\delta \in \mathcal{X}_\delta$ be chosen arbitrarily. Equation (7) follows from

$$\begin{aligned} \langle \dot{\varphi}_\delta, \bar{\varphi}_\delta \rangle_{\mathcal{H}} &= \sum_{k=1}^K \int_{I_k} \left\langle \frac{\varphi_\delta(t_k) - \varphi_\delta(t_{k-1})}{\Delta t_k}, \frac{\varphi_\delta(t_k) + \varphi_\delta(t_{k-1})}{2} \right\rangle_H dt \\ &= \frac{1}{2} \sum_{k=1}^K (\|\varphi_\delta(t_k)\|_H^2 - \|\varphi_\delta(t_{k-1})\|_H^2) = \frac{1}{2} \|\varphi_\delta(T)\|_H^2, \end{aligned}$$

where we have used that $\varphi_\delta(0) = 0$ in H .

To show (8) let $\varphi_\delta \in \mathcal{X}_\delta$ and $\phi_\delta \in \mathcal{Y}_\delta$ be chosen arbitrarily. Then, it follows that

$$\begin{aligned} \varphi_\delta(t) &= \frac{1}{\Delta t_k} ((t_k - t)\varphi_\delta(t_{k-1}) + (t - t_{k-1})\varphi_\delta(t_k)) && \text{f.a.a. } t \in I_k, \\ \bar{\varphi}_\delta(t) &= \frac{\varphi_\delta(t_{k-1}) + \varphi_\delta(t_k)}{2} && \text{f.a.a. } t \in I_k, \\ \phi_\delta(t) &= \phi_\delta^k(t) =: \phi_\delta^k && \text{f.a.a. } t \in I_k \end{aligned}$$

for every $k = 1, \dots, K$. Consequently,

$$\begin{aligned} \langle \varphi_\delta, \phi_\delta \rangle_{\mathcal{Y}} &= \frac{1}{\Delta t_k} \sum_{k=1}^K \left\langle \left[(t_k t - \frac{t^2}{2})\varphi_\delta(t_{k-1}) + (\frac{t^2}{2} - t_{k-1}t)\varphi_\delta(t_k), \phi_\delta^k \right]_{t=t_{k-1}}^{t=t_k} \right\rangle_V \\ &= \sum_{k=1}^K \frac{1}{\Delta t_k} \left\langle \frac{(t_k - t_{k-1})^2}{2} ((\varphi_\delta(t_{k-1}) + \varphi_\delta(t_k)), \phi_\delta^k \right\rangle_V \\ &= \sum_{k=1}^K \frac{\Delta t_k}{2} \langle \varphi_\delta(t_{k-1}) + \varphi_\delta(t_k), \phi_\delta^k \rangle_V \end{aligned}$$

and

$$\begin{aligned} \langle \bar{\varphi}_\delta, \phi_\delta \rangle_{\mathcal{Y}} &= \sum_{k=1}^K \int_{I_k} \left\langle \frac{\varphi_\delta(t_{k-1}) + \varphi_\delta(t_k)}{2}, \phi_\delta^k \right\rangle_V dt \\ &= \sum_{k=1}^K \frac{\Delta t_k}{2} \langle \varphi_\delta(t_{k-1}) + \varphi_\delta(t_k), \phi_\delta^k \rangle_V, \end{aligned}$$

which implies $\langle \varphi_\delta - \bar{\varphi}_\delta, \phi_\delta \rangle_{\mathcal{Y}} = 0$. Analogously, we derive that $\langle \varphi_\delta - \bar{\varphi}_\delta, \phi_\delta \rangle_{\mathcal{H}} = 0$. This finishes the proof. \square

2.3 RB space-time formulation

Analogously the space-time RB setting can be formulated and interpreted as a CN scheme. This CN interpretation guides us in the numerical calculation of solutions to the space-time formulation and in the derivation of an a-posteriori error estimator. For a spatial RB space $V_\ell = \text{span}\{\psi_1, \dots, \psi_\ell\} \subset V_h$ of dimension $\ell \in \mathbb{N}$, we introduce the RB solution and test spaces

$$\mathcal{X}_{\text{rb}} = S_{\Delta t} \otimes V_\ell, \quad \mathcal{Y}_{\text{rb}} = Q_{\Delta t} \otimes V_\ell,$$

respectively, where $\text{rb} = (\Delta t, \ell)$ stands for the temporal discretization and RB parameter. Note that it is also possible to consider RB spaces with respect to space and time, cf. [7, 8] for the linear FE and [32] for the linear RB case, but an adoption of the analysis presented in Chapter 3 is necessary.

For $\mu \in \mathcal{P}_{\text{ad}}$ we call $y_{\text{rb}} = y_{\text{rb}}(\mu) \in \mathcal{X}_{\text{rb}}$ an RB solution to (3) if

$$\mathcal{A}(y_{\text{rb}}, \phi; \mu) = \mathcal{F}(\phi; \mu) \quad \text{for all } \phi \in \mathcal{Y}_{\text{rb}}. \tag{9}$$

As previously done in the case of the FE space-time formulation, we can also reformulate (9) as a CN scheme, by using a trapezoidal quadrature rule. Therefore let $\Psi_\ell = [\psi_1, \dots, \psi_\ell] \in \mathbb{R}^{N \times \ell}$ denote the RB coefficient matrix whose columns are the FE coefficient vectors of the reduced basis functions. Let $M_\ell^{\text{space}} = \Psi_\ell^T M_h^{\text{space}} \Psi_\ell \in \mathbb{R}^{\ell \times \ell}$ and $V_\ell^{\text{space}} = \Psi_\ell^T V_h^{\text{space}} \Psi_\ell \in \mathbb{R}^{\ell \times \ell}$ denote the spatial RB mass matrix and spatial RB stiffness matrix, respectively. This leads to a sequence of root finding problems for $k = 1, \dots, K$:

$$G_{\text{rb}}^k(y_{\text{rb}}^k; \mu) = 0 \quad \text{in } \mathbb{R}^\ell$$

with

$$\begin{aligned} G_{\text{rb}}^k(y_{\text{rb}}^k; \mu) &= \frac{1}{\Delta t_k} M_\ell^{\text{space}} (y_{\text{rb}}^k - y_{\text{rb}}^{k-1}) + \frac{1}{2} [c(\mu) V_\ell^{\text{space}} (y_{\text{rb}}^k + y_{\text{rb}}^{k-1}) \\ &\quad + a(\mu) \Psi_\ell^T \tilde{M}_h^{\text{space}} (\max\{0, \Psi_\ell y_{\text{rb}}^k\} + \max\{0, \Psi_\ell y_{\text{rb}}^{k-1}\}) \\ &\quad - \Psi_\ell^T (F^k(\mu) + F^{k-1}(\mu))] \end{aligned}$$

with initial condition $y_{\text{rb}}^0 = 0$. Again this problem can be solved by applying a semismooth Newton method, where the k -th iteration matrix is given by

$$\begin{aligned} H_{\text{rb}}^k(y_{\text{rb}}^k; \mu) &= \frac{1}{\Delta t_k} M_\ell^{\text{space}} \\ &\quad + \frac{1}{2} \left(c(\mu) V_\ell^{\text{space}} + a(\mu) \Psi_\ell^T \tilde{M}_h^{\text{space}} \Theta(\Psi_\ell y_{\text{rb}}^k) \Psi_\ell \right) \in \mathbb{R}^{\ell \times \ell}. \end{aligned}$$

Since Ψ_ℓ has full rank, Remark 2 is applicable again, i.e. the sequence of roots y_{rb}^k is unique and every Newton iteration is uniquely solvable. Furthermore we obtain a stability estimate analogous to Proposition 1.

Corollary 3 *Let Assumption 1 hold. Then, for every $\mu \in \mathcal{P}_{ad}$, the RB solution $y_{rb}(\mu)$ to (9) satisfies the estimate*

$$\|\dot{y}_{rb}(\mu)\|_{\mathcal{Y}_\mathcal{I}} \leq \|f(\mu)\|_{\mathcal{Y}_\mathcal{I}}.$$

Proof The claim follows by similar arguments utilized to prove of Proposition 1. \square

3 Reduced basis method

In this section a greedy procedure for the generation of the RB space V_ℓ will be presented. Efficient error estimation is necessary and an error estimator is derived based on the space-time formulations introduced in Section 2. Compared to the elliptic case, cf. [25], the derived error estimator is composed of an additional term $\Delta_{\mathcal{P}^s}$. A large part of this section is dedicated to its convergence analysis, ultimately proving that $\Delta_{\mathcal{P}^s} = \mathcal{O}(\Delta t)$, where an equidistant temporal discretization is assumed for ease of presentation. To include a non-equidistant temporal discretization, we have to modify our discrete function spaces, but the convergence results remain valid with $\Delta t = \max_{k=1, \dots, K} t_k - t_{k-1}$.

3.1 Generation of reduced basis

For the generation of the spatial RB space V_ℓ proper orthogonal decomposition (POD) is used in a greedy procedure. For more details on POD we refer, e.g., to [33]. With $\mathcal{P}^\ell : V \rightarrow V_\ell$ the V -orthogonal projection onto the spatial RB space is denoted. By $\Delta(\mu)$ we denote an error estimator for the RB solution corresponding to the parameter $\mu \in \mathcal{P}$ with respect to RB solution space \mathcal{X}_{rb} and RB test space \mathcal{Y}_{rb} generated by V_ℓ . Furthermore POD_1 denotes the extraction of a dominant POD mode with respect to the inner product in V . The offline basis generation is summarized in Algorithm 1.

Algorithm 1 (POD-greedy RB method).

Require: Discrete training set $\mathcal{P}_{train} \subset \mathcal{P}_{ad}$, error tolerance $\varepsilon_{tol} > 0$;
 1: Set $\ell = 0, \Psi_0 = \emptyset, V_0 = \{0\}$;
 2: **while** $\varepsilon_\ell := \max\{\Delta(\mu) \mid \mu \in \mathcal{P}_{train}\} > \varepsilon_{tol}$ **do**
 3: Compute $\mu_{\ell+1} \in \arg \max\{\Delta(\mu) \mid \mu \in \mathcal{P}_{train}\}$;
 4: Set $e_{\ell+1}^k = y_\delta^k(\mu_{\ell+1}) - \mathcal{P}^\ell y_\delta^k(\mu_{\ell+1})$ for $k = 1, \dots, K$;
 5: Define $\mathcal{P}_{\ell+1} = \mathcal{P}_\ell \cup \{\mu_{\ell+1}\}$;
 6: Compute $\psi_{\ell+1} \in \text{POD}_1(\{e_{\ell+1}^k\}_{k=1}^K)$;
 7: Set $\Psi_{\ell+1} = \Psi_\ell \cup \{\psi_{\ell+1}\}, V_{\ell+1} = V_\ell \oplus \text{span}(\psi_{\ell+1})$ and $\ell = \ell + 1$;
 8: **end while**
 9: **return** Reduced basis Ψ_ℓ , spatial RB space V_ℓ .

It is possible to add more than just one POD mode per iteration to the reduced basis. This usually results in faster convergence of the algorithm and thus lower offline computational cost. But on the other hand it does not guarantee that the reduced basis is of minimal size. Note that the initial condition, which is zero, is always perfectly approximated.

3.2 RB error estimation

An efficient error estimator is crucial not only for Algorithm 1, but also for quality certification of the calculated RB solution in an online phase. Since $y_\delta, y_{rb} \notin \mathcal{Y}_{rb}$ we cannot test with the error $y_\delta - y_{rb}$ in the derivation of an error estimator. Therefore we will use an additional projection operator \mathcal{P}^δ that satisfies the following hypotheses.

Assumption 2 Let $\mathcal{P}^\delta : \mathcal{X}_\delta \rightarrow \mathcal{Y}_\delta$ be a bounded projection operator such that:

a) \mathcal{P}^δ satisfies

$$\langle \dot{\varphi}_\delta, \mathcal{P}^\delta \varphi_\delta \rangle_{\mathcal{J}_\mathcal{C}} = \frac{1}{2} \|\varphi_\delta(T)\|_H^2 \quad \text{for all } \varphi_\delta \in \mathcal{X}_\delta.$$

b) \mathcal{P}^δ satisfies

$$\langle \varphi_\delta - \mathcal{P}^\delta \varphi_\delta, \phi_\delta \rangle_{\mathcal{Y}} = \langle \varphi_\delta - \mathcal{P}^\delta \varphi_\delta, \phi_\delta \rangle_{\mathcal{J}_\mathcal{C}} = 0 \quad \text{for all } \varphi_\delta \in \mathcal{X}_\delta, \phi_\delta \in \mathcal{Y}_\delta.$$

We have already seen an example of a suitable projection operator in the previous section.

Example 1 For $1 \leq k \leq K$ and $t \in I_k$ almost everywhere we introduce the projection

$$(\mathcal{P}^\delta \varphi_\delta)(t) = \bar{\varphi}_\delta^k \quad \text{for all } \varphi_\delta \in \mathcal{X}_\delta,$$

where $\bar{\varphi}_\delta^k$ is defined in (5). Due to Lemma 2, Assumption 2 is satisfied. ◇

Before stating the main theorem of this section, we summarize properties of the projection operator \mathcal{P}^δ in the next lemma.

Lemma 4 Suppose that the projection operator \mathcal{P}^δ satisfies Assumption 2-b). For $\varphi_\delta \in \mathcal{X}_\delta$ it holds that

- a) $\|\varphi_\delta - \mathcal{P}^\delta \varphi_\delta\|_{\mathcal{J}_\mathcal{C}} \leq \|\varphi_\delta - \phi_\delta\|_{\mathcal{J}_\mathcal{C}}$ for all $\phi_\delta \in \mathcal{Y}_\delta$.
- b) $\|\varphi_\delta - \mathcal{P}^\delta \varphi_\delta\|_{\mathcal{J}_\mathcal{C}} \leq \Delta t \|\dot{\varphi}_\delta\|_{\mathcal{J}_\mathcal{C}}$, where $\Delta t = \max_{1 \leq k \leq K} \Delta t_k$ denotes the maximal time step.

Proof a) For $\varphi_\delta \in \mathcal{X}_\delta$ and $\phi_\delta \in \mathcal{Y}_\delta$ Assumption 2-b) and $\mathcal{P}^\delta \varphi_\delta \in \mathcal{Y}_\delta$ imply that

$$\begin{aligned} \|\varphi_\delta - \mathcal{P}^\delta \varphi_\delta\|_{\mathcal{J}_\mathcal{C}}^2 &= \langle \varphi_\delta - \mathcal{P}^\delta \varphi_\delta, \varphi_\delta \rangle_{\mathcal{J}_\mathcal{C}} - \langle \varphi_\delta - \mathcal{P}^\delta \varphi_\delta, \mathcal{P}^\delta \varphi_\delta \rangle_{\mathcal{J}_\mathcal{C}} \\ &= \langle \varphi_\delta - \mathcal{P}^\delta \varphi_\delta, \varphi_\delta \rangle_{\mathcal{J}_\mathcal{C}} \\ &= \langle \varphi_\delta - \mathcal{P}^\delta \varphi_\delta, \varphi_\delta - \phi_\delta \rangle_{\mathcal{J}_\mathcal{C}} \leq \|\varphi_\delta - \mathcal{P}^\delta \varphi_\delta\|_{\mathcal{J}_\mathcal{C}} \|\varphi_\delta - \phi_\delta\|_{\mathcal{J}_\mathcal{C}}, \end{aligned}$$

which gives the claim.

- b) Denote by $\Pi^\delta: \mathcal{X}_\delta \rightarrow \mathcal{Y}_\delta$ the piecewise constant interpolation operator defined as $(\Pi^\delta \varphi_\delta)(t) = \varphi_\delta(t_{k-1})$ for $1 \leq k \leq K$ such that $t \in I_k$. For $e(t) = \varphi_\delta(t) - (\Pi^\delta \varphi_\delta)(t) \in V_h$, we denote by e_k the restriction of e on the interval I_k . Then e_k is affine linear w.r.t. t and belongs to $H^1(I_k; V)$. Thus, for $t \in I_k$ we estimate

$$e(t) = e_k(t) = \underbrace{e_k(t_{k-1})}_{=0} + \int_{t_{k-1}}^t \dot{e}_k(s) \, ds$$

by the fundamental theorem of calculus. This implies

$$\begin{aligned} \|e_k(t)\|_H &= \left\| \int_{t_{k-1}}^t \dot{e}_k(s) \, ds \right\|_H \leq \int_{I_k} \|\dot{e}_k(s)\|_H \, ds \\ &\leq \left(\int_{I_k} 1 \, ds \right)^{1/2} \left(\int_{I_k} \|\dot{e}_k(s)\|_H^2 \, ds \right)^{1/2} = \sqrt{\Delta t_k} \|\dot{e}_k\|_{L^2(I_k; H)}. \end{aligned}$$

Consequently,

$$\|e_k\|_{L^2(I_k; H)}^2 = \int_{I_k} \|e_k(s)\|_H^2 \, ds \leq (\Delta t_k)^2 \|\dot{e}_k\|_{L^2(I_k; H)}^2.$$

Since $\dot{e}_k = \dot{\varphi}_\delta$ on I_k this implies that

$$\|e_k\|_{L^2(I_k; H)} \leq \Delta t_k \|\dot{\varphi}_\delta\|_{L^2(I_k; H)}.$$

We have $\Pi^\delta \varphi_\delta \in \mathcal{Y}_\delta$ and $\Delta t = \max_{1 \leq k \leq K} \Delta t_k$. Hence, using part a) it follows

$$\begin{aligned} \|\varphi_\delta - \mathcal{P}^\delta \varphi_\delta\|_{\mathcal{J}_C}^2 &\leq \|\varphi_\delta - \Pi^\delta \varphi_\delta\|_{\mathcal{J}_C}^2 = \sum_{k=1}^K \|\varphi_\delta - \Pi^\delta \varphi_\delta\|_{L^2(I_k; H)}^2 \\ &\leq \sum_{k=1}^K (\Delta t_k)^2 \|\dot{\varphi}_\delta\|_{L^2(I_k; H)}^2 \leq (\Delta t)^2 \|\dot{\varphi}_\delta\|_{\mathcal{J}_C}^2. \end{aligned}$$

This implies part b). □

Next we define the RB residual as

$$\mathcal{R}_{\text{rb}}(\phi; \mu) = \mathcal{F}(\phi; \mu) - \mathcal{A}(y_{\text{rb}}(\mu), \phi; \mu) \quad \text{for all } \phi \in \mathcal{Y}_\delta. \tag{10}$$

Now we are ready to state the RB error estimator.

Theorem 5 *Let Assumption 1 hold. Suppose that the projection operator \mathcal{P}^δ satisfies Assumption 2. For $\mu \in \mathcal{P}_{ad}$ denote the RB error as $e_{rb}(\mu) = y_\delta(\mu) - y_{rb}(\mu)$ satisfying $\mathcal{P}^\delta e_{rb}(\mu) \neq 0$. Then,*

$$\|\mathcal{P}^\delta e_{rb}(\mu)\|_{\mathcal{Y}} \leq \Delta_{rb}(\mu) + \Delta_{\mathcal{P}^\delta}(\mu), \tag{11}$$

with RB error estimator Δ_{rb} and projection error estimator $\Delta_{\mathcal{P}^\delta}$ given by

$$\begin{aligned} \Delta_{rb}(\mu) &= \frac{1}{c(\mu)} \|\mathcal{R}_{rb}(\cdot; \mu)\|_{\mathcal{Y}'_\delta}, \\ \Delta_{\mathcal{P}^\delta}(\mu) &= \frac{a(\mu)}{c(\mu)} \frac{\|\max\{0, y_\delta(\mu)\} - \max\{0, y_{rb}(\mu)\}\|_{\mathcal{H}_\mathcal{C}}}{\|\mathcal{P}^\delta e_{rb}(\mu)\|_{\mathcal{Y}}} \|e_{rb}(\mu) - \mathcal{P}^\delta e_{rb}(\mu)\|_{\mathcal{H}_\mathcal{C}}, \end{aligned}$$

respectively.

Proof Let $\mu \in \mathcal{P}_{ad}$ be chosen arbitrarily. From $e_{rb}(\mu) \in \mathcal{X}_\delta$ we obtain by Assumption 2-b), the monotonicity of max and Assumption 2-a) that

$$\begin{aligned} \|\mathcal{P}^\delta e_{rb}(\mu)\|_{\mathcal{Y}}^2 &= \langle \mathcal{P}^\delta e_{rb}(\mu), \mathcal{P}^\delta e_{rb}(\mu) \rangle_{\mathcal{Y}} = \langle e_{rb}(\mu), \mathcal{P}^\delta e_{rb}(\mu) \rangle_{\mathcal{Y}}, \\ a(\mu) \langle \max\{0, y_\delta(\mu)\} - \max\{0, y_{rb}(\mu)\}, e_{rb}(\mu) \rangle_{\mathcal{H}_\mathcal{C}} &\geq 0, \\ \langle \dot{e}_{rb}(\mu), \mathcal{P}^\delta e_{rb}(\mu) \rangle_{\mathcal{H}_\mathcal{C}} &= \langle \dot{e}_{rb}(\mu), e_{rb}(\mu) \rangle_{\mathcal{H}_\mathcal{C}} = \frac{1}{2} \|e_{rb}(T; \mu)\|_H^2 \geq 0. \end{aligned}$$

Thus we infer that

$$\begin{aligned} c(\mu) \|\mathcal{P}^\delta e_{rb}(\mu)\|_{\mathcal{Y}}^2 &= c(\mu) \langle e_{rb}(\mu), \mathcal{P}^\delta e_{rb}(\mu) \rangle_{\mathcal{Y}} \\ &\leq c(\mu) \langle e_{rb}(\mu), \mathcal{P}^\delta e_{rb}(\mu) \rangle_{\mathcal{Y}} + a(\mu) \langle \max\{0, y_\delta(\mu)\} - \max\{0, y_{rb}(\mu)\}, e_{rb}(\mu) \rangle_{\mathcal{H}_\mathcal{C}} \\ &\leq c(\mu) \langle e_{rb}(\mu), \mathcal{P}^\delta e_{rb}(\mu) \rangle_{\mathcal{Y}} + \langle \dot{e}_{rb}(\mu), \mathcal{P}^\delta e_{rb}(\mu) \rangle_{\mathcal{H}_\mathcal{C}} \\ &\quad + a(\mu) (\langle \max\{0, y_\delta(\mu)\} - \max\{0, y_{rb}(\mu)\}, \mathcal{P}^\delta e_{rb}(\mu) \rangle_{\mathcal{H}_\mathcal{C}} \\ &\quad + \langle \max\{0, y_\delta(\mu)\} - \max\{0, y_{rb}(\mu)\}, e_{rb}(\mu) - \mathcal{P}^\delta e_{rb}(\mu) \rangle_{\mathcal{H}_\mathcal{C}}). \end{aligned}$$

Utilizing (1) it follows that

$$\begin{aligned} c(\mu) \|\mathcal{P}^\delta e_{rb}(\mu)\|_{\mathcal{Y}}^2 &\leq a(\mu) \langle \max\{0, y_\delta(\mu)\} - \max\{0, y_{rb}(\mu)\}, e_{rb}(\mu) - \mathcal{P}^\delta e_{rb}(\mu) \rangle_{\mathcal{H}_\mathcal{C}} \\ &\quad + \mathcal{A}(y_\delta(\mu), \mathcal{P}^\delta e_{rb}(\mu); \mu) - \mathcal{A}(y_{rb}(\mu), \mathcal{P}^\delta e_{rb}(\mu); \mu). \end{aligned}$$

Moreover, we deduce from (3) that

$$\mathcal{A}(y_\delta(\mu), \mathcal{P}^\delta e_{rb}(\mu); \mu) = \mathcal{F}(\mathcal{P}^\delta e_{rb}(\mu); \mu).$$

Consequently, by using (10)

$$\begin{aligned}
 c(\mu) \|\mathcal{P}^\delta e_{rb}(\mu)\|_{\mathcal{Y}}^2 &\leq \mathcal{R}_{rb}(\mathcal{P}^\delta e_{rb}(\mu); \mu) \\
 &+ a(\mu) \langle \max\{0, y_\delta(\mu)\} - \max\{0, y_{rb}(\mu)\}, e_{rb}(\mu) - \mathcal{P}^\delta e_{rb}(\mu) \rangle_{\mathcal{H}} \\
 &\leq \|\mathcal{R}_{rb}(\cdot; \mu)\|_{\mathcal{Y}'_\delta} \|\mathcal{P}^\delta e_{rb}(\mu)\|_{\mathcal{Y}} \\
 &+ a(\mu) \|\max\{0, y_\delta(\mu)\} - \max\{0, y_{rb}(\mu)\}\|_{\mathcal{H}} \|e_{rb}(\mu) - \mathcal{P}^\delta e_{rb}(\mu)\|_{\mathcal{H}}.
 \end{aligned}$$

Thus we obtain

$$\begin{aligned}
 \|\mathcal{P}^\delta e_{rb}(\mu)\|_{\mathcal{Y}} &\leq \frac{1}{c(\mu)} \|\mathcal{R}_{rb}(\cdot; \mu)\|_{\mathcal{Y}'_\delta} \\
 &+ \frac{a(\mu)}{c(\mu)} \frac{\|\max\{0, y_\delta(\mu)\} - \max\{0, y_{rb}(\mu)\}\|_{\mathcal{H}} \|e_{rb}(\mu) - \mathcal{P}^\delta e_{rb}(\mu)\|_{\mathcal{H}}}{\|\mathcal{P}^\delta e_{rb}(\mu)\|_{\mathcal{Y}}}
 \end{aligned}$$

which gives the claim. □

Finally we study the convergence of the projection error estimator $\Delta_{\mathcal{P}^\delta}(\mu)$.

Proposition 6 *Let Assumptions 1 and 2 hold. Denote $\Delta t = \max_{1 \leq k \leq K} \Delta t_k$. Then $\Delta_{\mathcal{P}^\delta}(\mu) \rightarrow 0$ for $\Delta t \rightarrow 0$ and every $\mu \in \mathcal{P}_{ad}$. Especially for sufficiently small Δt and every $\mu \in \mathcal{P}_{ad}$ the estimate*

$$\Delta_{\mathcal{P}^\delta}(\mu) \leq C \Delta t$$

with a constant $C \geq 0$ independent of μ is satisfied.

Proof Without loss of generality, we can assume that $e_{rb}(\mu) \neq 0$ for the parameters μ under consideration, or there is nothing to show. Recall that $e_{rb}(\mu) = y_\delta(\mu) - y_{rb}(\mu) \in \mathcal{X}_\delta$ holds true. By Lemma 4-b), Proposition 1 and Corollary 3 we obtain

$$\begin{aligned}
 \|e_{rb}(\mu) - \mathcal{P}^\delta e_{rb}(\mu)\|_{\mathcal{H}} &\leq \|y_\delta(\mu) - \mathcal{P}^\delta y_\delta(\mu)\|_{\mathcal{H}} + \|y_{rb}(\mu) - \mathcal{P}^\delta y_{rb}(\mu)\|_{\mathcal{H}} \\
 &\leq (\|\dot{y}_\delta(\mu)\|_{\mathcal{H}} + \|\dot{y}_{rb}(\mu)\|_{\mathcal{H}}) \Delta t \leq 2 \|f(\mu)\|_{\mathcal{H}} \Delta t
 \end{aligned} \tag{12}$$

for every $\mu \in \mathcal{P}_{ad}$. Thus $\mathcal{P}^\delta e_{rb}(\mu) \rightarrow e_{rb}(\mu)$ in \mathcal{H} for $\Delta t \rightarrow 0$. This implies that there exists a (sufficiently small) constant $\tau_\mu \in (0, T]$ satisfying

$$0 \leq \frac{\|e_{rb}(\mu)\|_{\mathcal{H}}}{\|\mathcal{P}^\delta e_{rb}(\mu)\|_{\mathcal{H}}} \leq \frac{3}{2} \quad \text{for } \Delta t \in (0, \tau_\mu]. \tag{13}$$

Here we have used that without loss of generality $\mathcal{P}^\delta e_{rb}(\mu) \neq 0$ for all $\Delta t \in (0, \tau_\mu]$, possibly after shrinking τ_μ , because $\mathcal{P}^\delta e_{rb}(\mu) \rightarrow e_{rb}(\mu)$ in \mathcal{H} for $\Delta t \rightarrow 0$ and $e_{rb}(\mu) \neq 0$ holds. From Poincaré’s inequality we infer that there exists a constant

$c_P > 0$ so that

$$\frac{1}{\|\varphi\|_{\mathcal{Y}}} \leq \frac{c_P}{\|\varphi\|_{\mathcal{H}}} \quad \text{for all } \varphi \in \mathcal{Y} \setminus \{0\}. \tag{14}$$

Utilizing (12)-(14) we obtain for $\Delta t \in (0, \tilde{C}(\mu)]$

$$\begin{aligned} \Delta_{\mathcal{P}^\delta}(\mu) &= \frac{a(\mu)}{c(\mu)} \frac{\|e_{rb}(\mu)\|_{\mathcal{H}}}{\|\mathcal{P}^\delta e_{rb}(\mu)\|_{\mathcal{Y}}} \|e_{rb}(\mu) - \mathcal{P}^\delta e_{rb}(\mu)\|_{\mathcal{H}} \\ &\leq \frac{a(\mu)}{c(\mu)} \frac{c_P}{\|\mathcal{P}^\delta e_{rb}(\mu)\|_{\mathcal{H}}} (2\|f(\mu)\|_{\mathcal{H}} \Delta t) \leq 3c_P \frac{a(\mu)}{c(\mu)} \|f(\mu)\|_{\mathcal{H}} \Delta t \leq \hat{C}(\mu) \Delta t, \end{aligned}$$

where the nonnegative constant $\hat{C}(\mu) = 3a(\mu)c_P \|f(\mu)\|_{\mathcal{H}}/c(\mu)$ is uniformly bounded w.r.t. μ due to the Lipschitz continuity of a, c, f and the compactness of \mathcal{P}_{ad} . □

4 Adaptive RB-DEIM method

The current model order reduction approach is unsatisfactory for two reasons. First, the sequence of root finding problems $G_{rb}^k(y_{rb}^k; \mu)$ is on spatial RB level ℓ , but an evaluation of the nonsmoothness on spatial FE level N is necessary. Furthermore the dual norm of the residual for the RB error estimator $\Delta_{rb}(\mu)$ cannot be efficiently computed in an offline-online separable fashion. This is typical for RB error estimation in the context of nonlinear PDEs, cf. [11, 25]. We will use the discrete empirical interpolation method (DEIM) to approximate the nonsmoothness and overcome those difficulties. For details on the efficient evaluation of the dual norm of the residual in combination with (D)EIM, we refer to [34, Section 4.2.5].

In the first half of this section the classical DEIM framework is presented together with a new approach for error estimation. Afterwards an adaptive DEIM framework that combines RB and DEIM offline phases is presented.

4.1 Classical DEIM

We keep the presentation of the classical DEIM approach short and refer to [20, 21] for more details on the general procedure and to [31] for an application to an elliptic max-type PDE. For the remainder of this section we will assume that the right-hand sides F^k are parameter separable independently of the time instance, i.e.,

$$F^k(\mu) = B_F \beta_F(\mu) \gamma_F(k),$$

with

$$B_F \in \mathbb{R}^{N \times P}, \quad \beta_F: \mathcal{P} \rightarrow \mathbb{R}^P, \quad \gamma_F: \{0, \dots, K\} \rightarrow \mathbb{R}.$$

If this should not be the case, a further DEIM approximation would be necessary. The procedure can be easily generalized to this situation, cf. [20]. Now DEIM introduces a projection matrix $P = [e_{i_1} | \dots | e_{i_L}] \in \mathbb{R}^{N \times L}$ and an approximation matrix $\Phi = [\phi_1 | \dots | \phi_L] \in \mathbb{R}^{N \times L}$. We assume $L \ll N$ and that $P^T \Phi \in \mathbb{R}^{L \times L}$ is invertible. Then the max term in G_{rb}^k can be approximated as

$$\Phi(P^T \Phi)^{-1} \max \{0, P^T \Psi_\ell y_{rb}^k(\mu)\}.$$

This gives the RB-DEIM approximation of the RB root finding problem presented in Section 2.3. For $k = 1, \dots, K$ solve the sequence of root finding problems:

$$G_{rb,L}^k(y_{rb,L}^k; \mu) = 0 \quad \text{in } \mathbb{R}^\ell \tag{15}$$

with

$$\begin{aligned} G_{rb,L}^k(y_{rb,L}^k; \mu) = & \frac{1}{\Delta t_k} M_\ell^{\text{space}}(y_{rb,L}^k - y_{rb,L}^{k-1}) + \frac{1}{2} \left[c(\mu) V_\ell^{\text{space}}(y_{rb,L}^k + y_{rb,L}^{k-1}) \right. \\ & + a(\mu) \Psi_\ell^T \tilde{M}_h^{\text{space}} \Phi(P^T \Phi)^{-1} (\max \{0, P^T \Psi_\ell y_{rb,L}^k\}) \\ & \left. + \max \{0, P^T \Psi_\ell y_{rb,L}^{k-1}\} \right] - \Psi_\ell^T (F^k(\mu) + F^{k-1}(\mu)), \end{aligned}$$

with initial condition $y_{rb,L}^0 = 0$. Again this problem can be solved by applying a semismooth Newton method, where the k -th iteration matrix is given by

$$\begin{aligned} H_{rb,L}^k(y_{rb,L}^k; \mu) = & \frac{1}{\Delta t_k} M_\ell^{\text{space}} \\ & + \frac{1}{2} (c(\mu) V_h^{\text{space}} + a(\mu) \Psi_\ell^T \tilde{M}_h^{\text{space}} \Phi(P^T \Phi)^{-1} \Theta(P^T \Psi_\ell y_{rb,L}^k) P^T \Psi_\ell) \in \mathbb{R}^{\ell \times \ell} \end{aligned}$$

with $\Theta(v) = \max\{0, v\} \in \mathbb{R}^L$ for $v \in \mathbb{R}^L$. Problem (15) is now independent of the spatial FE dimension N . Note that we cannot guarantee the uniqueness of $y_{rb,L}^k$ and the unique solvability of the Newton iteration by the same arguments as for the FE and RB problem. In practice, a sufficiently accurate DEIM approximation of the nonlinearity will also restore the monotonicity of the max term and thus the monotonicity argument from Remark 2 can be applied again, cf. [11] for an analogous argumentation in the context of EIM. We will only briefly comment on how P and Φ are generated:

- 1) A discrete DEIM training set $\mathcal{P}_{\text{train}}^L \subset \mathcal{P}_{\text{ad}}$ is chosen;
- 2) Snapshots are generated as $\max\{0, y_\delta^k(\mu)\}$ for $\mu \in \mathcal{P}_{\text{train}}^L$ and $k = 1, \dots, L$;
- 3) The DEIM algorithm generates P and Φ from the snapshots.

4.2 RB-DEIM error estimation

Often the additional error to the solution due to DEIM is not included in the error estimator. The problem is that DEIM approximates the nonlinearity on a discrete level. Therefore working with the DEIM approximation in the variational formulation is problematic. It is also problematic to efficiently incorporate the DEIM error later, since this would mean that also a Riesz representative must be calculated for the additional error quantity, which, to the best of our knowledge, is not possible in a parameter separable fashion. Thus we want to present an idea on how to already (partially) discretize the residual and incorporate the DEIM error early.

For an element $\varphi \in \mathcal{Y}_\delta \cup \mathcal{X}_\delta$ in the FE solution or test space, we denote the nodal values at time $t \in I$ by $\hat{\varphi}(t) \in \mathbb{R}^N$. Now we introduce the RB-DEIM residual:

$$\mathcal{R}_{\text{rb},L}(\phi; \mu) = \mathcal{F}(\phi; \mu) - \mathcal{A}_L(y_{\text{rb},L}(\mu), \phi; \mu) \quad \text{for all } \phi \in \mathcal{Y}_\delta \quad (16)$$

with

$$\begin{aligned} \mathcal{A}_L(y_{\text{rb},L}(\mu), \phi; \mu) &= \mathcal{B}(y_{\text{rb},L}(\mu), \phi; \mu) + \mathcal{N}_L(y_{\text{rb},L}(\mu), \phi; \mu), \\ \mathcal{N}_L(y_{\text{rb},L}(\mu), \phi; \mu) &= a(\mu) \int_I \langle \Phi(\mathbf{P}^\top \Phi)^{-1} \max\{0, \mathbf{P}^\top \hat{y}_{\text{rb},L}(\mu)\}, \hat{\phi} \rangle_{\tilde{\mathbf{M}}_h^{\text{space}}} dt \end{aligned}$$

and $\langle \cdot, \cdot \rangle_{\tilde{\mathbf{M}}_h^{\text{space}}} = \langle \tilde{\mathbf{M}}_h^{\text{space}} \cdot, \cdot \rangle_2$, where $\langle \cdot, \cdot \rangle_2$ denotes the Euclidean norm. Moreover, \mathcal{N}_L represents the approximation obtained by mass lumping and DEIM.

Remark 3 The sequence of root finding problems $G_{\text{rb},L}^k$ can equivalently be derived by working with \mathcal{Y}_{rb} as test space in (16) and applying a trapezoidal quadrature rule analogously to Section 2.2. This further justifies the idea of already incorporating DEIM in the variational formulation. \diamond

The idea is now to proceed analogously to Theorem 5. Before we state the main result of this section, let us introduce the notations $\hat{\mathcal{H}} = L^2(I; \mathbf{M}_h^{\text{space}})$ and $\tilde{\mathcal{H}} = L^2(I; \tilde{\mathbf{M}}_h^{\text{space}})$, where $L^2(I; \mathbf{M}_h^{\text{space}})$ and $L^2(I; \tilde{\mathbf{M}}_h^{\text{space}})$ denote the spaces of all (measurable) functions $\varphi: I \rightarrow \mathbb{R}^N$ satisfying

$$\begin{aligned} \|\varphi\|_{\hat{\mathcal{H}}} &= \left(\int_I \|\varphi(t)\|_{\mathbf{M}_h^{\text{space}}}^2 dt \right)^{1/2} < \infty, \\ \|\varphi\|_{\tilde{\mathcal{H}}} &= \left(\int_I \|\varphi(t)\|_{\tilde{\mathbf{M}}_h^{\text{space}}}^2 dt \right)^{1/2} < \infty, \end{aligned}$$

respectively. Note that

$$\|\varphi\|_{\tilde{\mathcal{H}}}^2 = \|\hat{\varphi}\|_{\hat{\mathcal{H}}}^2 \leq \|\hat{\varphi}\|_{\tilde{\mathcal{H}}}^2 + \epsilon_{\tilde{\mathcal{H}}}(\varphi, \varphi),$$

with the mass lumping error defined as

$$\epsilon_{\mathcal{J}_C}(\varphi, \psi) = |\langle \varphi, \psi \rangle_{\mathcal{J}_C} - \langle \hat{\varphi}, \hat{\psi} \rangle_{\hat{\mathcal{J}}_C}| \quad \text{for all } \varphi, \psi \in \mathcal{Y}_\delta \cup \mathcal{X}_\delta. \tag{17}$$

Proposition 7 *Let Assumption 1 hold. Suppose that the projection operator \mathcal{P}^δ satisfies Assumption 2. For $\mu \in \mathcal{P}_{ad}$ denote the RB-DEIM error as $e_{rb,L}(\mu) = y_\delta(\mu) - y_{rb,L}(\mu)$ satisfying $\mathcal{P}^\delta e_{rb,L}(\mu) \neq 0$. Then,*

$$\|e_{rb,L}(\mu)\|_{\mathcal{Y}} \leq \Delta_{rb,L}(\mu) + \Delta_{\mathcal{P}^\delta}(\mu) + \Delta_L(\mu) + \Delta_{\hat{\mathcal{J}}_C}(\mu),$$

with DEIM approximated RB error estimator $\Delta_{rb,L}$, projection error estimator $\Delta_{\mathcal{P}^\delta}$, DEIM error estimator Δ_L and mass lumping error estimator $\Delta_{\hat{\mathcal{J}}_C}$ given by

$$\begin{aligned} \Delta_{rb,L}(\mu) &= \frac{1}{c(\mu)} \|\mathcal{R}_{rb,L}(\cdot; \mu)\|_{\mathcal{Y}'_\delta}, \\ \Delta_{\mathcal{P}^\delta}(\mu) &= \frac{a(\mu)}{c(\mu)} \frac{\|\max\{0, y_\delta(\mu)\} - \max\{0, y_{rb,L}(\mu)\}\|_{\mathcal{J}_C}}{\|\mathcal{P}^\delta e_{rb,L}(\mu)\|_{\mathcal{Y}}} \|e_{rb,L}(\mu) - \mathcal{P}^\delta e_{rb,L}(\mu)\|_{\mathcal{J}_C}, \\ \Delta_L(\mu) &= \frac{cPa(\mu)}{c(\mu)} \|\Phi(P^\top \Phi)^{-1} \max\{0, P^\top \hat{y}_{rb,L}(\mu)\} - \max\{0, \hat{y}_{rb,L}(\mu)\}\|_{\hat{\mathcal{J}}_C}, \\ \Delta_{\hat{\mathcal{J}}_C}(\mu) &= \frac{a(\mu)}{c(\mu)\|\mathcal{P}^\delta e_{rb,L}(\mu)\|_{\mathcal{Y}}} \left[\epsilon_{\hat{\mathcal{J}}_C}(\max\{0, y_{rb,L}(\mu)\}, \mathcal{P}^\delta e_{rb,L}(\mu)) \right. \\ &\quad \left. + \|\Phi(P^\top \Phi)^{-1} \max\{0, P^\top \hat{y}_{rb,L}(\mu)\} - \max\{0, \hat{y}_{rb,L}(\mu)\}\|_{\hat{\mathcal{J}}_C} \right. \\ &\quad \left. \cdot \epsilon_{\hat{\mathcal{J}}_C}(\mathcal{P}^\delta e_{rb,L}(\mu), \mathcal{P}^\delta e_{rb,L}(\mu)) \right]. \end{aligned}$$

Proof Let $\mu \in \mathcal{P}_{ad}$ be fixed. Analogously to the proof of Theorem 5 we obtain

$$\begin{aligned} &c(\mu) \|\mathcal{P}^\delta e_{rb,L}(\mu)\|_{\mathcal{Y}}^2 \\ &\leq \mathcal{R}_{rb,L}(\mathcal{P}^\delta e_{rb,L}(\mu); \mu) \\ &\quad + a(\mu) \left[-\langle \max\{0, y_{rb,L}(\mu), \mathcal{P}^\delta e_{rb,L}(\mu)\} \rangle_{\mathcal{J}_C} \right. \\ &\quad \quad + \langle \max\{0, y_\delta(\mu)\} - \max\{0, y_{rb,L}(\mu)\}, e_{rb,L}(\mu) - \mathcal{P}^\delta e_{rb,L}(\mu) \rangle_{\mathcal{J}_C} \\ &\quad \quad \left. + \langle \Phi(P^\top \Phi)^{-1} \max\{0, P^\top \hat{y}_{rb,L}(\mu)\}, \mathcal{P}^\delta \hat{e}_{rb,L}(\mu) \rangle_{\hat{\mathcal{J}}_C} \right] \\ &\leq \|\mathcal{R}_{rb,L}(\cdot; \mu)\|_{\mathcal{Y}'_\delta} \|\mathcal{P}^\delta e_{rb,L}(\mu)\|_{\mathcal{Y}} \\ &\quad + a(\mu) \|\max\{0, y_\delta(\mu)\} - \max\{0, y_{rb,L}(\mu)\}\|_{\mathcal{J}_C} \|e_{rb,L}(\mu) - \mathcal{P}^\delta e_{rb,L}(\mu)\|_{\mathcal{J}_C} \\ &\quad + a(\mu) \left| \langle \Phi(P^\top \Phi)^{-1} \max\{0, P^\top \hat{y}_{rb,L}(\mu)\}, \mathcal{P}^\delta \hat{e}_{rb,L}(\mu) \rangle_{\hat{\mathcal{J}}_C} \right. \\ &\quad \quad \left. - \langle \max\{0, y_{rb,L}(\mu), \mathcal{P}^\delta e_{rb,L}(\mu)\} \rangle_{\mathcal{J}_C} \right|. \end{aligned}$$

Now we use the mass lumping error defined in (17) to estimate

$$\begin{aligned}
 & a(\mu) \left| \langle \Phi(\mathbf{P}^\top \Phi)^{-1} \max\{0, \mathbf{P}^\top \hat{y}_{rb,L}(\mu)\}, \mathcal{P}^\delta \hat{e}_{rb,L}(\mu) \rangle_{\mathcal{H}} \right. \\
 & \quad \left. - \langle \max\{0, y_{rb,L}(\mu)\}, \mathcal{P}^\delta e_{rb,L}(\mu) \rangle_{\mathcal{H}} \right| \\
 & \leq a(\mu) \left| \langle \Phi(\mathbf{P}^\top \Phi)^{-1} \max\{0, \mathbf{P}^\top \hat{y}_{rb,L}(\mu)\}, \mathcal{P}^\delta \hat{e}_{rb,L}(\mu) \rangle_{\mathcal{H}} \right. \\
 & \quad \left. - \langle \max\{0, \hat{y}_{rb,L}(\mu)\}, \mathcal{P}^\delta \hat{e}_{rb,L}(\mu) \rangle_{\mathcal{H}} + \epsilon_{\mathcal{H}}(\max\{0, y_{rb,L}(\mu)\}, \mathcal{P}^\delta e_{\ell,L}(\mu)) \right| \\
 & \leq a(\mu) \left[\|\Phi(\mathbf{P}^\top \Phi)^{-1} \max\{0, \mathbf{P}^\top \hat{y}_{rb,L}(\mu)\} - \max\{0, \hat{y}_{rb,L}(\mu)\}\|_{\mathcal{H}} \|\mathcal{P}^\delta \hat{e}_{rb,L}(\mu)\|_{\mathcal{H}} \right. \\
 & \quad \left. + \epsilon_{\mathcal{H}}(\max\{0, y_{rb,L}(\mu)\}, \mathcal{P}^\delta e_{rb,L}(\mu)) \right] \\
 & \leq a(\mu) \left[\|\Phi(\mathbf{P}^\top \Phi)^{-1} \max\{0, \mathbf{P}^\top \hat{y}_{rb,L}(\mu)\} - \max\{0, \hat{y}_{rb,L}(\mu)\}\|_{\mathcal{H}} \right. \\
 & \quad \cdot [c_P \|\mathcal{P}^\delta e_{rb,L}(\mu)\|_{\mathcal{Y}} + \epsilon_{\mathcal{H}}(\mathcal{P}^\delta e_{rb,L}(\mu), \mathcal{P}^\delta e_{rb,L}(\mu))] \\
 & \quad \left. + \epsilon_{\mathcal{H}}(\max\{0, y_{rb,L}(\mu)\}, \mathcal{P}^\delta e_{rb,L}(\mu)) \right],
 \end{aligned}$$

where we have used

$$\|\mathcal{P}^\delta \hat{e}_{rb,L}(\mu)\|_{\mathcal{H}} \leq \|\mathcal{P}^\delta \hat{e}_{rb,L}(\mu)\|_{\mathcal{H}} + \epsilon_{\mathcal{H}}(\mathcal{P}^\delta e_{rb,L}(\mu), \mathcal{P}^\delta e_{rb,L}(\mu))$$

and Poincarés inequality with constant $c_P > 0$ in the last estimate. All in all we obtain

$$\begin{aligned}
 & \|\mathcal{P}^\delta e_{rb,L}(\mu)\|_{\mathcal{Y}} \leq \frac{1}{c(\mu)} \|\mathcal{R}_{rb,L}(\cdot; \mu)\|_{\mathcal{Y}'_s} \\
 & + \frac{a(\mu)}{c(\mu)} \frac{\|\max\{0, y_s(\mu)\} - \max\{0, y_{rb,L}(\mu)\}\|_{\mathcal{H}}}{\|\mathcal{P}^\delta e_{rb,L}(\mu)\|_{\mathcal{Y}}} \|e_{rb,L}(\mu) - \mathcal{P}^\delta e_{rb,L}(\mu)\|_{\mathcal{H}} \\
 & + \frac{a(\mu)c_P}{c(\mu)} \|\Phi(\mathbf{P}^\top \Phi)^{-1} \max\{0, \mathbf{P}^\top \hat{y}_{rb,L}(\mu)\} - \max\{0, \hat{y}_{rb,L}(\mu)\}\|_{\mathcal{H}} \\
 & + \frac{a(\mu)}{c(\mu)\|\mathcal{P}^\delta e_{rb,L}(\mu)\|_{\mathcal{Y}}} \left[\|\Phi(\mathbf{P}^\top \Phi)^{-1} \max\{0, \mathbf{P}^\top \hat{y}_{\ell,L}(\mu)\} - \max\{0, \hat{y}_{\ell,L}(\mu)\}\|_{\mathcal{H}} \right. \\
 & \quad \left. \cdot \epsilon_{\mathcal{H}}(\mathcal{P}^\delta e_{\ell,L}(\mu), \mathcal{P}^\delta e_{\ell,L}(\mu)) + \epsilon_{\mathcal{H}}(\max\{0, y_{rb,L}(\mu)\}, \mathcal{P}^\delta e_{rb,L}(\mu)) \right]
 \end{aligned}$$

All in all the desired estimate follows. □

Remark 4 a) Note that analogously to $\Delta_{\mathcal{P}^\delta}$ also $\Delta_{\mathcal{H}}^{\mathcal{H}}$ can be made small independently of the chosen RB space, by choosing a sufficiently rich spatial FE space. This follows as a consequence of [35, Lemma 1] adopted to the present PDE. Therefore we will use $\Delta_{rb,L} + \Delta_L$ as error estimator in our numerical experiments.

- b) The computation of Δ_L requires an evaluation of the nonsmoothness on FE level. This cannot be avoided, but is computationally still cheap compared to the other computations involved in the basis generation process, which especially include the computation of FE solutions for basis updates and the application of POD. \diamond

4.3 Adaptive DEIM

In Proposition 7 a decomposition of the total error into a temporal mesh-dependent part $\Delta_{\mathcal{P}^s}$, a spatial mesh-dependent part $\Delta_{\tilde{\mathcal{J}}_C}$, an RB dependent part $\Delta_{\text{rb},L}$ and a DEIM dependent part Δ_L has been proposed. As already discussed, we will neglect the mesh-dependent parts, since they can be chosen arbitrarily small by sufficiently rich spatial and temporal FE spaces. Now typically a DEIM basis would be generated in an offline phase to efficiently approximate the nonsmoothness in a lower dimensional subspace as described in Section 4.1. This process is decoupled from the RB basis generation described in Section 3.1. More recently ideas have been presented on how to combine RB and DEIM basis generation into one adaptive process, see e.g. [36, 37]. Advantages of such ideas are control over the DEIM error in the offline phase and thus a DEIM basis of suitable size. Also the combination of RB and DEIM basis generation leads to fewer computational costs. The adaptive DEIM algorithm in [37] is tailored towards optimization and updates the DEIM basis without increasing its size. This results in only locally good approximation quality, which in optimization is sufficient. In [36] the authors essentially calculate a completely new DEIM basis whenever it needs to be updated. On one hand, this is cheap, since no additional snapshot generation is necessary and the size of the singular value decomposition necessary for DEIM is usually still small compared to classical offline DEIM. On the other hand, this usually means that all preassembled DEIM data must be thrown away whenever the DEIM basis is updated. Therefore we suggest a successive DEIM approach for the space-time setting, which might result in a (slightly) larger DEIM basis, but allows to reuse previous information. We summarize this approach in Algorithm 2.

To avoid stagnation at the end of the algorithm one should e.g. choose $\varepsilon_{\text{tol}}^{\text{rb}} = 0.1\varepsilon_{\text{tol}}$ and $\varepsilon_{\text{tol}}^L = 0.01\varepsilon_{\text{tol}}$. This ensures that the algorithm aims at individual tolerances $\varepsilon_{\text{tol}}^{\text{rb}}$ and $\varepsilon_{\text{tol}}^L$ well below ε_{tol} . Also it should be ensured that the DEIM basis is always at least as large as the RB basis for numerical stability. This is the reason why there is the possibility to start the algorithm with an initial DEIM basis. For more details we refer to [36]. The algorithm performs best, if the construction of a RB and a DEIM basis are of equal difficulty, i.e. if both RB and DEIM error estimator contribute equally to the overall error estimator. This is often not the case and usually the construction of a DEIM basis is more difficult, as also our numerical experiments in Section 5 suggest, see the right plot in Figs. 2 and 3. Heuristics to let the DEIM basis grow faster exist, e.g. \log_{10} could be replaced by \log_2 in line 11. This can be very problem specific and is beyond the scope of this work. In our implementation a maximum of one element is added to the RB basis per iteration. In [36] a logarithmic term analogue to the one for the DEIM basis is used. In our opinion this might lead to too many elements being added during the first iterations and wrong elements removed later on. But of course the algorithm could also be extended in this fashion.

Algorithm 2 (Adaptive POD-greedy RB-DEIM method).

Require: Initial DEIM data P_0, Φ_0, l_0 , discrete training set of parameters $\mathcal{P}_{\text{train}} \subset \mathcal{P}_{\text{ad}}$, error tolerances $\varepsilon_{\text{tol}}^{\text{rb}}, \varepsilon_{\text{tol}}^L > 0$ and $\varepsilon_{\text{tol}} \geq \varepsilon_{\text{tol}}^{\text{rb}} + \varepsilon_{\text{tol}}^L$;

- 1: Set $\ell = 0, \Psi_0 = \emptyset, V_0 = \{0\}, P = P_0, \Phi = \Phi_0$ and $l = l_0$;
- 2: Compute $\tilde{\mu} \in \arg \max\{\Delta_{\text{rb},L}(\mu) + \Delta_L(\mu) \mid \mu \in \mathcal{P}_{\text{train}}\}$;
- 3: Set $\varepsilon_1 = \Delta_{\text{rb},L}(\tilde{\mu})$ and $\varepsilon_2 = \Delta_L(\tilde{\mu})$;
- 4: **while** $\varepsilon_1 + \varepsilon_2 > \varepsilon_{\text{tol}}$ **do**
- 5: % RB part;
- 6: **if** $\varepsilon_1 > \varepsilon_{\text{tol}}^{\text{rb}}$ **then**
- 7: Set $e_{\ell+1}^k = y_{\delta}^k(\tilde{\mu}) - \mathcal{P}_{V_{\ell}} y_{\delta}^k(\tilde{\mu})$ for $k = 1, \dots, K$;
- 8: Compute $\psi_{\ell+1} \in \text{POD}_1(\{e_{\ell+1}^k\}_{k=1}^K)$;
- 9: Set $\Psi_{\ell+1} = \Psi_{\ell} \cup \{\psi_{\ell+1}\}, V_{\ell+1} = V_{\ell} \oplus \text{span}(\psi_{\ell+1})$ and $\ell = \ell + 1$;
- 10: **end if**
- 11: % DEIM part;
- 12: Set $L = \lfloor \log_{10}(\varepsilon_2/\varepsilon_{\text{tol}}^L) \rfloor$;
- 13: **if** $L = 0$ **then**
- 14: $L = 1$ (avoid stagnation);
- 15: **end if**
- 16: Set $E = [\max\{0, y_{\delta}^1(\tilde{\mu}), \dots, \max\{0, y_{\delta}^K(\tilde{\mu})\}]$; % snapshot matrix
- 17: Put $E = E - \Phi \Phi^{\top} E$; % subtract previous DEIM information
- 18: Set $\tilde{L} = \min\{L, \text{rank}(E)\}$;
- 19: Compute POD basis $\tilde{\Phi}$ of length \tilde{L} w.r.t. identity matrix for E ;
- 20: **if** $\Phi = []$ (no initial DEIM basis) **then**
- 21: Compute $i \in \arg \max_{k=1, \dots, N} |(\tilde{\Phi}_1)_k|$;
- 22: Set $\Phi = [\tilde{\Phi}_1], l = [i], P = [e_i]$ and $\tilde{j} = 2$;
- 23: **else**
- 24: Set $\tilde{j} = 1$;
- 25: **end if**
- 26: **for** $j = \tilde{j}, \dots, \tilde{L}$ **do**
- 27: Solve $(P^{\top} \Phi) \gamma = P^{\top} \tilde{\Phi}_j$;
- 28: Set $r = \tilde{\Phi}_j - \Phi \gamma$;
- 29: Compute $i \in \arg \max_{k=1, \dots, N} |r_k|$;
- 30: Set $\Phi = [\Phi, \tilde{\Phi}_j], l = [l, i]$ and $P = [P, e_i]$;
- 31: **end for**
- 32: % Error update;
- 33: Compute $\tilde{\mu} \in \arg \max\{\Delta_{\text{rb},L}(\mu) + \Delta_L(\mu) \mid \mu \in \mathcal{P}_{\text{train}}\}$;
- 34: Set $\varepsilon_1 = \Delta_{\text{rb},L}(\tilde{\mu})$ and $\varepsilon_L = \Delta_L(\tilde{\mu})$;
- 35: **end while**
- 36: **return** Reduced basis Ψ_{ℓ} , spatial RB space V_{ℓ} , DEIM data P, Φ, l .

5 Numerical experiments

In this section we present two numerical examples to investigate the performance of the different space-time RB approaches. Parameter choices fixed for both examples are summarized in Table 1.

The adaptive DEIM algorithm uses an initial DEIM basis of size two generated from one snapshot. Regarding the settings of the examples, let us mention that the discretization of the parameter space is going to be a trade-off between a sufficient fineness and the necessary computational time. Furthermore the fineness of the discretized

Table 1 Fixed parameters for the numerical examples

PDE (Section 2.1)	Discretization (Section 2.2)		Tolerances (Algorithm 2)		
Ω	K	$1/h$	ε_{tol}	$\varepsilon_{\text{tol}}^{\text{rb}}$	$\varepsilon_{\text{tol}}^L$
$(0, 1)^2$	400	{50, 100, 200}	10^{-3}	10^{-4}	10^{-5}

Recall that K and h denote the number of time grid points and the spatial step-size, respectively

parameter spaces is $1/3$ for both examples. The functions a and c are chosen such that the nonsmoothness is stronger for parameters with larger norm. Note that if the nonsmoothness is too strong, its approximation becomes problematic, e.g. increasing a by a factor of 100 will lead to the algorithms struggling, especially with the DEIM approximation.

Our code is implemented in Python 3 and uses FEniCS (see [38]) for the matrix assembly. Sparse memory management and computations are implemented with SciPy (see [39]). All computations below were run on an Ubuntu 20.04 notebook with 32 GB main memory and an Intel Core i7-8565U CPU.

5.1 Example 1

We choose $T = 20$ for the maximal time horizon and set $\mathcal{P}_{\text{ad}} = [-10, 10]$, $a(\mu) = 1 + 2|\mu|$ and $c(\mu) = 5/(5 + |\mu|)$. Furthermore the right-hand sides are

$$f(t; \mu)(x_1, x_2) = \underbrace{10 \sin\left(\frac{4\pi t}{T}\right)}_{\text{time}} \underbrace{\sqrt{1+t} \left(\frac{1}{2} - x_1\right) \sin(\pi x_1) \sin(\pi x_2)}_{\text{space}} \underbrace{\mu}_{\text{parameter}}$$

For the classical RB and RB-DEIM approach we fix a training set of 60 equidistant parameters and a test set of 100 equidistant parameters. For DEIM we additionally choose the same training set as for the RB offline phase and choose the same parameter L we obtain from the adaptive RB-DEIM. Note that this means that already information from the adaptive approach is used in the classical RB-DEIM approach, but it allows for a better comparison of the result, which are shown in Table 2.

First of all, we can observe mesh independence in the average efficiency, which is the arithmetic mean of the quotient of error estimator and true error, and the size of the RB basis for the approaches without DEIM. If DEIM is used, the results are slightly mesh dependent, due to the mesh dependency of DEIM. Also the average error and error estimator are in the same regions and far below the tolerance of 10^{-3} for all approaches and independently of the mesh. We can observe that the RB error estimator leads to slight decreases in offline computational time, but only the introduction of DEIM allows for efficient evaluations of the estimator. Unfortunately the offline computational cost of DEIM outweighs this advantage. Finally we can observe that the adaptive RB-DEIM approach leads to the largest reduction in offline computational

Table 2 Example 1

$1/h$	Av. sp.-up	Time offline	Av. error	Av. est.	Av. proj. error	Av. eff.	Size RB	Size DEIM
RB (true error)								
50	5.95	$7.76 \cdot 10^2$	$2.73 \cdot 10^{-4}$	$2.88 \cdot 10^{-4}$	$4.73 \cdot 10^{-6}$	1.04	15	–
100	39.54	$4.54 \cdot 10^3$	$5.14 \cdot 10^{-4}$	$5.17 \cdot 10^{-4}$	$1.72 \cdot 10^{-6}$	1.00	16	–
200	93.72	$2.03 \cdot 10^4$	$5.62 \cdot 10^{-4}$	$5.64 \cdot 10^{-4}$	$1.44 \cdot 10^{-6}$	1.00	15	–
RB (estimator)								
50	5.48	$9.39 \cdot 10^2$	$2.72 \cdot 10^{-4}$	$2.88 \cdot 10^{-4}$	$4.74 \cdot 10^{-6}$	1.04	15	–
100	39.66	$3.54 \cdot 10^3$	$6.62 \cdot 10^{-4}$	$6.65 \cdot 10^{-4}$	$1.51 \cdot 10^{-6}$	1.00	15	–
200	91.17	$1.32 \cdot 10^4$	$6.21 \cdot 10^{-4}$	$6.23 \cdot 10^{-4}$	$1.34 \cdot 10^{-6}$	1.00	15	–
RB-DEIM (estimator)								
50	8.06	$3.07 \cdot 10^2 + 5.67 \cdot 10^2$	$1.86 \cdot 10^{-4}$	$3.22 \cdot 10^{-4}$	$1.56 \cdot 10^{-6}$	1.59	18	72
100	94.29	$3.51 \cdot 10^3 + 1.86 \cdot 10^3$	$1.78 \cdot 10^{-4}$	$3.80 \cdot 10^{-4}$	$3.28 \cdot 10^{-7}$	3.37	27	66
200	879.23	$1.82 \cdot 10^4 + 6.29 \cdot 10^3$	$4.28 \cdot 10^{-4}$	$6.19 \cdot 10^{-4}$	$3.54 \cdot 10^{-7}$	1.45	19	87
RB-DEIM (estimator, adaptive)								
50	7.68	$6.43 \cdot 10^2$	$1.66 \cdot 10^{-4}$	$2.93 \cdot 10^{-4}$	$8.90 \cdot 10^{-7}$	1.84	20	72
100	96.84	$1.67 \cdot 10^3$	$2.65 \cdot 10^{-4}$	$5.10 \cdot 10^{-4}$	$1.11 \cdot 10^{-6}$	2.43	19	66
200	882.92	$6.62 \cdot 10^3$	$2.72 \cdot 10^{-4}$	$5.10 \cdot 10^{-4}$	$9.77 \cdot 10^{-7}$	2.33	19	87

Comparison of RB and RB-DEIM approaches for different spatial discretizations $1/h$. For RB-DEIM (estimator) offline time is given as DEIM + RB time

cost, whilst simultaneously giving equivalent benefits as classical RB-DEIM in average speed-up in the online phase. Especially on finer grids the advantages of an RB-DEIM approach are evident. Last but not least, we want to mention two noticeable entries in Table 2. First, the RB basis without estimator contains one more element than that with estimator for $1/h = 100$. This is atypical, but of course possible due to the outstanding efficiency and the fact that the greedy procedure is only locally choosing the best update in every iteration. Second, the RB-DEIM basis without adaptivity is rather large and contains 27 elements for $1/h = 100$. This comes from the fact that the procedure struggles to reach the desired overall tolerance in this case.

To further investigate the error estimator, it is shown in Fig. 1 together with the error on the test set.

We can see that the RB error estimator clearly mimics the behavior of the true error. In the RB-DEIM approach, the additional DEIM error estimator is more volatile. Furthermore the additional DEIM error estimator is necessary to avoid underestimation, though this is hard to observe in the plot.

Finally we want to investigate the projection error and the evolution of the different error quantities during the adaptive RB-DEIM basis generation. The average projection error, on the previously introduced test set, is shown in Fig. 2 on the left and the evolution of the errors during training is shown on the right.

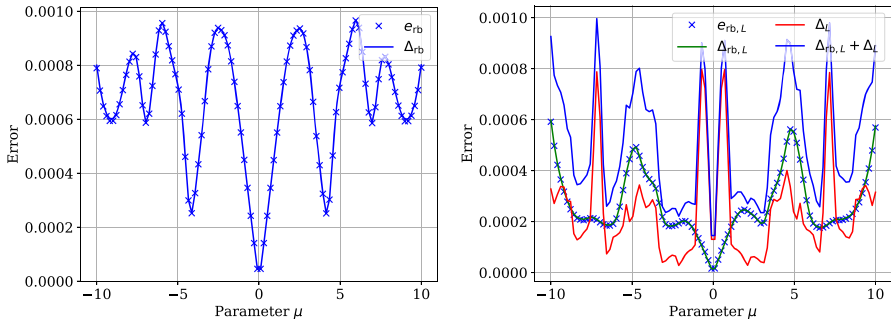


Fig. 1 Example 1. Error and RB error estimator (left) and error and RB-DEIM error estimator (right) with $1/h = 100$

In the left plot we observe that the projection error $\Delta_{\mathcal{P}^\delta}$ decreases with increasing time grid points as proposed in Proposition 6. However, a linear convergence can only be expected asymptotically. In particular, the regime of linear convergence seems to not be reached yet. This can be expected due to the quotient in the projection error estimator, which is only asymptotically one. Nonetheless the values of the projection error estimator are very small and can be neglected as suggested. In the right plot we can see that the RB error tends to be lower than the DEIM error, but both decrease sufficiently fast.

All in all this example clearly suggests the usage of the RB error estimator over the calculation of the true error and the usage of the adaptive DEIM approach over the classical DEIM approach. We also investigated the projection error estimator and have shown that its small values justify neglectation.

5.2 Example 2

In the second numerical example we choose $T = 10$ for the maximal time horizon and use a two-dimensional parameter space with $\mathcal{P}_{\text{ad}} = [-2, 2]^2$, $a(\mu) = 1 + 5 \|\mu\|_2$

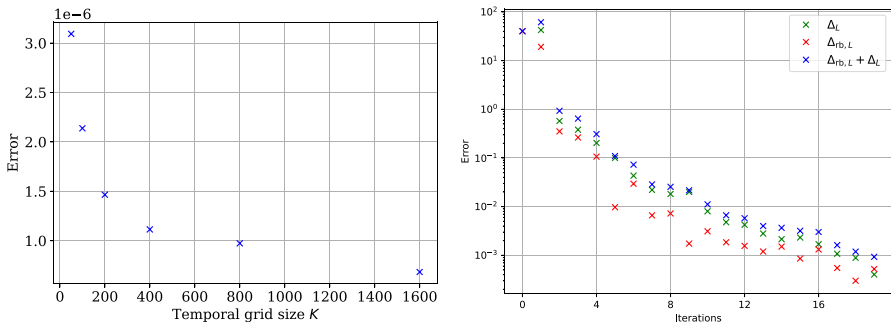


Fig. 2 Example 1. Left: Projection error $\Delta_{\mathcal{P}^\delta}$ for varying temporal grid size K . Right: Evolution of the total error $\Delta_{\text{rb},L} + \Delta_L$ and its individual components during training of the adaptive RB-DEIM. Both with $1/h = 100$

and $c(\mu) = 3/(1 + |\mu_1|)$. Furthermore, the right-hand sides $f(t; \mu)$ are chosen to be

$$f(t, \mathbf{x}; \mu) = \underbrace{10 \sin\left(\frac{4\pi t}{T}\right)}_{\text{time}} \underbrace{\sqrt{1+t} \begin{cases} x_1 x_2 \mu_1, & \text{for } x_1 \leq \frac{1}{2}, \\ x_1^2 x_2^2 \mu_2, & \text{otherwise.} \end{cases}}_{\text{space and parameter}}$$

This time we only present results for RB with error estimator and the adaptive RB-DEIM approach, since based on the theoretical results and Example 1, these are the most promising strategies. For RB we fix a training set of 144 equidistant parameters and a test set of 225 equidistant parameters. The results for the different spatial discretizations are shown in Table 3

Again the RB results show mesh independence in the average error, error estimator and efficiency and the size of the RB basis. This time the average efficiency for RB-DEIM is significantly larger (approx. 10 compared to approx. 1.01) which is one of the reasons for the increased DEIM basis. Note that the RB basis is also significantly larger (approx. 30) than for the RB approach (13), but both bases are mesh independent in size. This is a typical behavior that occurs since for the given example DEIM approximation is more difficult than RB approximation. Thus the RB basis is chosen in a way that a tolerance of 10^{-4} is reached, so that the DEIM tolerance only needs to be below $\varepsilon_{\text{tol}} = 10^{-3}$. Furthermore the size of the DEIM basis is not mesh independent, but slightly increases for finer meshes.

Again we want to investigate the projection error and the evolution of the different error quantities during the adaptive RB-DEIM basis generation. The average projection error is shown in Fig. 3 on the left and the evolution of the errors during training is shown on the right.

Table 3 Example 2

$1/h$	Av. sp.-up	Time offline	Av. error	Av. est.	Av. proj. error	Av. eff.	Size RB	Size DEIM
RB (estimator)								
50	5.40	$2.08 \cdot 10^3$	$3.09 \cdot 10^{-4}$	$3.14 \cdot 10^{-4}$	$5.66 \cdot 10^{-7}$	1.01	13	–
100	45.79	$5.98 \cdot 10^3$	$3.16 \cdot 10^{-4}$	$3.22 \cdot 10^{-4}$	$6.03 \cdot 10^{-7}$	1.01	13	–
200	96.04	$2.82 \cdot 10^4$	$3.15 \cdot 10^{-4}$	$3.21 \cdot 10^{-4}$	$6.02 \cdot 10^{-7}$	1.01	13	–
RB-DEIM (estimator, adaptive)								
50	7.58	$3.23 \cdot 10^3$	$7.25 \cdot 10^{-5}$	$1.04 \cdot 10^{-3}$	$1.96 \cdot 10^{-7}$	10.00	29	177
100	127.06	$7.97 \cdot 10^3$	$7.04 \cdot 10^{-5}$	$8.42 \cdot 10^{-4}$	$1.93 \cdot 10^{-7}$	9.63	30	212
200	858.93	$2.02 \cdot 10^4$	$5.78 \cdot 10^{-5}$	$8.11 \cdot 10^{-4}$	$2.25 \cdot 10^{-7}$	10.70	29	222

Comparison of RB and RB-DEIM approaches for different spatial discretizations $1/h$. For RB-DEIM (estimator) offline time is given as DEIM + RB time

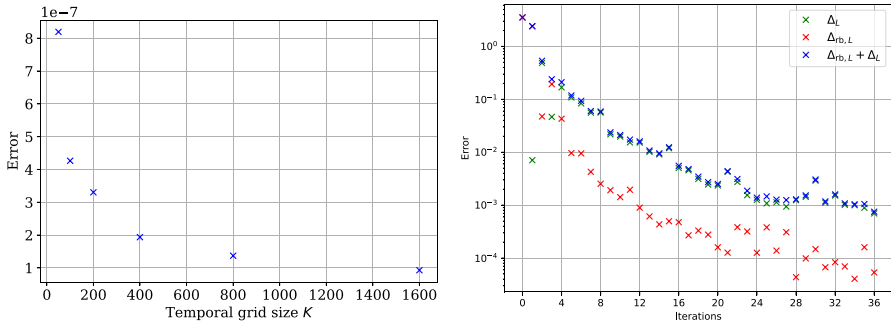


Fig. 3 Example 2. Left: Projection error $\Delta_{\mathcal{P}}$ for varying temporal grid size K . Right: Evolution of the total error $\Delta_{rb,L} + \Delta_L$ and its individual components during training of the adaptive RB-DEIM. Both with $1/h = 100$

In the left plot, we can observe that the average projection error is again neglectable, due to its small values and we can observe the same convergence behavior as discussed in detail for the previous example. In the right plot, we can observe that in contrast to Example 1, the slow convergence of the DEIM error slows down the overall convergence and leads to an RB error below $\varepsilon_{tol}^{rb} = 10^{-4}$. Therefore the DEIM error in essence only needs to be below $\varepsilon_{tol} = 10^{-3}$.

The second example again suggests the usage of the RB error estimator over the calculation of the true error, due to the excellent average efficiency. Again it is unproblematic to neglect the projection error estimator. Note that we have also excluded the RB-DEIM approach without adaptivity, since no convergence can be observed with the size L of the DEIM basis taken from the adaptive approach. This suggests advantages of the adaptive approach, but we can also observe that the adaptive algorithm begins to struggle when the DEIM approximation is significantly harder than the RB approximation, which is especially linked to the choices of the functions a and c .

6 Conclusion

We have introduced a novel space-time a-posteriori error estimator for a nonsmooth parabolic PDE. The numerical results show promising speed-up and good efficiency. Nonetheless the model order reduction is limited by the evaluation of the nonsmooth max-term. To solve this problem we have introduced a novel adaptive RB-DEIM approach based on a modified version of the space-time a-posteriori error estimator, which suggests that the total error estimator can be decomposed into a RB and a DEIM part. Again numerical results show the capabilities, but also possible limitations of this approach. As long as the DEIM approximations complexity is in the same regime as the RB approximations complexity, the novel approach works well. Especially the speed-up compared to RB approaches can be significantly increased, whilst maintaining the same approximation quality. Compared to classical RB-DEIM, this is only possible due to the adaptive nature of our algorithm.

Acknowledgements We are grateful to Denis Korolev (Berlin), Dominik Meidner (Munich) and Karsten Urban (Ulm) for fruitful discussions and very helpful remarks.

Funding Open Access funding enabled and organized by Projekt DEAL. This research was supported by the German Research Foundation (DFG) under grant number VO 1658/5-2 within the DFG Priority Program “Non-smooth and Complementarity-based Distributed Parameter Systems: Simulation and Hierarchical Optimization” (SPP 1962).

Declarations

Conflicts of interest The authors declare that there are no conflicts of interest regarding the publication of this paper.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Bernreuther, M.: Nonsmooth PDEs: efficient algorithms, model order reduction, multiobjective PDE-constrained optimization. PhD thesis, Universität Konstanz, Konstanz (2023). <http://nbn-resolving.de/urn:nbn:de:bsz:352-2-1jxe38e3x59983>
2. Rappaz, J.: Approximation of a nondifferentiable nonlinear problem related to MHD equilibria. *Numer. Math.* **45**(1), 117–133 (1984)
3. Xin, J.: An introduction to fronts in random media. *Surveys and Tutorials in the Applied Mathematical Sciences*, vol. 5. Springer, New York (2009)
4. Meidner, D., Vexler, B.: A priori error analysis of the Petrov-Galerkin Crank-Nicolson scheme for parabolic optimal control problems. *SIAM J. Control and Optimization*. **49**, 2183–2211 (2011). <https://doi.org/10.1137/100809611>
5. Gunzburger, M., Kunoth, A.: Space-time adaptive wavelet methods for optimal control problems constrained by parabolic evolution equations. *J. Control Optim.* **55**, 1150–1170 (2011). <https://doi.org/10.1137/100806382>
6. Neitzel, I., Vexler, B.: A priori error estimates for space-time finite element discretization of semilinear parabolic optimal control problems. *Numerische Mathematik*. **120**, 345–386 (2012). <https://doi.org/10.1007/s00211-011-0409-9>
7. Langer, U., Steinbach, O.: Space time methods: applications to partial differential equations. *Radon Series on Computational and Applied Mathematics*, vol. 25. De Gruyter, Berlin (2019)
8. Steinbach, O.: Space-time finite element methods for parabolic problems. *Comput. Methods Appl. Math.* **15**, 551–566 (2015). <https://doi.org/10.1515/cmam-2015-0026>
9. Steinbach, O., Yang, H.: Comparison of algebraic multigrid methods for an adaptive space-time finite element discretization of the heat equation in 3d and 4d. *Numer Linear Algebra Appl* **25**, 2143 (2018). <https://doi.org/10.1002/nla.2143>
10. Harbrecht, H., Tausch, J.: A fast sparse grid based space-time boundary element method for the nonstationary heat equation. *Numer. Math.* **140**, 239–264 (2018). <https://doi.org/10.1007/s00211-018-0963-5>
11. Hinze, M., Korolev, D.: A space-time certified reduced basis method for quasilinear parabolic partial differential equations. *Adv. Comput. Math.* **47**, 36 (2021). <https://doi.org/10.1007/s10444-021-09860-z>

12. Steih, K., Urban, K.: Space-time reduced basis methods for time-periodic partial differential equations. *IFAC Proc. Vol.* **45**, 710–715 (2012). <https://doi.org/10.3182/20120215-3-AT-3016.00126>
13. Yano, M., Patera, A.T., Urban, K.: A space-time hp-interpolation-based certified reduced basis method for Burgers' equation. *Math. Models Methods Appl. Sci.* **24**, 1903–1935 (2014). <https://doi.org/10.1142/S0218202514500110>
14. Urban, K., Patera, A.T.: An improved error bound for reduced basis approximation of linear parabolic problems. *Math. Comput.* **83**, 1599–1615 (2014). <https://doi.org/10.1090/S0025-5718-2013-02782-2>
15. 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. Model. Numer. Anal.* **56**, 1173–1198 (2022). <https://doi.org/10.1051/m2an/2022035>
16. Beranek, N., Reinhold, A., Urban, K.: A space-time variational method for optimal control problems: well-posedness, stability and numerical solution. Submitted (2022). <https://doi.org/10.48550/arXiv.2010.00345>
17. Ballarin, F., Rozza, G., Strazzullo, M.: Chapter 9 – Space-time POD-Galerkin approach for parametric flow control. In: Trélat, E., Zuazua, E. (eds.) *Numerical Control: Part A. Handbook of Numerical Analysis*, vol. 23, pp. 307–338. Elsevier, ??? (2022). <https://doi.org/10.1016/bs.hna.2021.12.009>
18. Strazzullo, M., Ballarin, F., Rozza, G.: A certified reduced basis method for linear parametrized parabolic optimal control problems in space-time formulation. Submitted (2021). <https://doi.org/10.48550/arXiv.2103.00460>
19. Barrault, M., Maday, Y., Nguyen, N.C., Patera, A.T.: An empirical interpolation method: application to efficient reduced-basis discretization of partial differential equations. *Comptes Rendus Mathématique.* **339**, 667–672 (2004). <https://doi.org/10.1016/j.crma.2004.08.006>
20. Chaturantabut, S., Sorensen, D.C.: Nonlinear model reduction via discrete empirical interpolation. *SIAM J. Sci. Comput.* **32**, 2737–2764 (2010). <https://doi.org/10.1137/090766498>
21. Chaturantabut, S., Sorensen, D.C.: A state space estimate for POD-DEIM nonlinear model reduction. *SIAM J. Numer. Anal.* **50**, 46–63 (2012). <https://doi.org/10.1137/110822724>
22. Betz, L.M.: Second-order sufficient optimality conditions for optimal control of non-smooth, semilinear parabolic equations. *J. Control. Optim.* **57**, 4033–4062 (2019). <https://doi.org/10.1137/19M1239106>
23. Meyer, C., Susu, L.M.: Optimal control of nonsmooth, semilinear parabolic equations. *J. Control. Optim.* **55**, 2206–2234 (2017). <https://doi.org/10.1137/15M1040426>
24. Drohmann, M., Haasdonk, B., Ohlberger, M.: Adaptive reduced basis methods for nonlinear convection-diffusion equations. In: *Finite Volumes for Complex Applications VI Problems & Perspectives* (2010). https://doi.org/10.1007/978-3-642-20671-9_39
25. Bernreuther, M., Müller, G., Volkwein, S.: Reduced basis model order reduction in optimal control of a nonsmooth semilinear elliptic PDE. In: Herzog, R., Heinkenschloss, M., Kalise, D., Stadler, G., Trélat, E. (eds.) *Optimization and Control for Partial Differential Equations*, pp. 1–32. De Gruyter, Berlin, Boston (2022). <https://doi.org/10.1515/9783110695984-001>
26. Evans, L.C.: *Partial differential equations*. Graduate Studies in Mathematics. American Mathematical Society, Providence, Rhode Island (2010)
27. Zeidler, E.: *Nonlinear functional analysis and its applications*. Linear Monotone Operators vol. II/A. Springer, New York (1989)
28. Zeidler, E.: *Nonlinear functional analysis and its applications*. Nonlinear Monotone Operators vol. II/B. Springer, New York (1989)
29. Thomée, V.: *Galerkin finite element methods for parabolic problems*. Springer, Berlin (1997)
30. Hintermüller, M.: *Semismooth Newton methods and applications*. Oberwolfach-Seminar on Mathematics of PDE-Constrained Optimization at Mathematisches Forschungsinstitut in Oberwolfach (2010). https://www.math.uni-hamburg.de/home/hinze/Psfiles/Hintermueller_OWNotes.pdf
31. Bernreuther, M.: *RB-based PDE-constrained non-smooth optimization*. Master's thesis, Universität Konstanz (2019). <http://nbn-resolving.de/urn:nbn:de:bsz:352-2-t4k1djjy77yn3>
32. Braun, J.: *Space-time reduced basis method for solving parameterized heat equations*. Master's thesis, Universität Konstanz (2023)
33. Gubisch, M., Volkwein, S.: Chapter 1: POD for linear-quadratic optimal control. In: *Model Reduction and Approximation - Theory and Algorithms*. Computational Science & Engineering, pp. 3–63. SIAM, Philadelphia (2017). <https://doi.org/10.1137/1.9781611974829.ch1>
34. Hesthaven, J.S., Rozza, G., Stamm, B.: *Certified reduced basis methods for parametrized partial differential equations*. SpringerBriefs in Mathematics. Springer, Cham (2016)

35. Zeng, J.-p., Yu, H.-x.: Error estimates of the lumped mass finite element method for semilinear elliptic problems. *Journal of Computational and Applied Mathematics*. **236**, 1993–2004 (2012). <https://doi.org/10.1016/j.cam.2011.11.009>
36. Chellappa, S., Feng, L., Benner, P.: Adaptive basis construction and improved error estimation for parametric nonlinear dynamical systems. *Int. J. Numer. Methods Eng.* **121**, 5320–5349 (2020). <https://doi.org/10.1002/nme.6462>
37. Peherstorfer, B., Willcox, K.: Online adaptive model reduction for nonlinear systems via low-rank updates. *SIAM J. Sci. Comput.* **37**, 2123–2150 (2015). <https://doi.org/10.1137/140989169>
38. Alnas, M., Blechta, J., Hake, J., Johansson, A., Kehlet, B., Logg, A., Richardson, C., Ring, J., Rognes, M.E., Wells, G.N.: The FEniCS project version 1.5. *Archive of Numerical Software*. **3** (2015). <https://doi.org/10.11588/ans.2015.100.20553>
39. Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S.J., Brett, M., Wilson, J., Jarrod Millman, K., Mayorov, N., Nelson, A.R.J., Jones, E., Kern, R., Larson, E., Carey, C.J., Polat, İ., Feng, Y., Moore, E.W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E.A., Harris, C.R., Archibald, A.M., Ribeiro, A.H., Pedregosa, F., van Mulbregt, P., Contributors, S...: SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods*. **17**, 261–272 (2020). <https://doi.org/10.1038/s41592-019-0686-2>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.