DFG Deutsche Forschungsgemeinschaft
Priority Programme 1962

Reduced Basis Model Order Reduction in Optimal Control of a Nonsmooth Semilinear Elliptic PDE

Marco Bernreuther, Georg Müller, Stefan Volkwein

SPP 1962 | Non-smooth and Complementarity-based Distributed Parameter Systems: Simulation and Hierarchical Optimization

Preprint Number SPP1962-138

received on April 18, 2020

Edited by
SPP1962 at Weierstrass Institute for Applied Analysis and Stochastics (WIAS)
Leibniz Institute in the Forschungsverbund Berlin e.V.
Mohrenstraße 39, 10117 Berlin, Germany
E-Mail: spp1962@wias-berlin.de

World Wide Web: http://spp1962.wias-berlin.de/

Marco Bernreuther*, Georg Müller*, and Stefan Volkwein*

Reduced Basis Model Order Reduction in Optimal Control of a Nonsmooth Semilinear Elliptic PDE

Abstract: In this paper, an optimization problem governed by a nonsmooth semilinear elliptic partial differential equation is considered. A reduced order approach is applied in order to obtain a computationally fast and certified numerical solution approach. Using the reduced basis method and efficient a-posteriori error estimation for the primal and dual equations, an adaptive algorithm is developed and tested successfully for several numerical examples.

Keywords: Nonsmooth optimization, nonsmooth semilinear elliptic equations, semismooth Newton, reduced basis method, error estimation, strong stationarity.

1 Introduction

In this paper, we consider the optimal control problem governed by a nonsmooth semilinear elliptic partial differential equation (PDE)

$$\begin{aligned} &\min_{(y,\mu)} \mathcal{J}(y,\mu) = j(y) + \frac{\sigma}{2}\,\|\mu\|_{\mathsf{A}}^2 \\ &\text{s.t. } (y,\mu) \in V \times \mathbb{R}^p \text{ satisfies } -\Delta y + \max\{0,y\} = \mathcal{B}\mu \text{ in } V', \end{aligned} \tag{$\mathbf{P}$}$$

where μ denotes a parameter that acts as a control on the right hand side and y is the state. We endow $V \coloneqq H_0^1(\Omega)$ with the usual inner product

$$\langle \varphi, \phi \rangle_V \coloneqq \int_\Omega \nabla\varphi \cdot \nabla\phi + \varphi\phi \,\mathrm{d}\boldsymbol{x} \quad \text{for } \varphi, \phi \in V$$

and the induced norm $\|\cdot\|_V \coloneqq \langle \cdot\,, \cdot \rangle_V^{1/2}$. Its topological dual space is written as V'. Our assumptions on the data throughout the paper are as follows:

***Corresponding author: Marco Bernreuther, Georg Müller, Stefan Volkwein,** University of Konstanz, Department of Mathematics and Statistics, WG Numerical Optimization, Universitätsstraße 10, 78457 Konstanz, Germany, e-mail: Stefan.Volkwein@uni-konstanz.de

Assumption 1.1. 1) $\Omega \subset \mathbb{R}^d$, $d \geq 1$, *is a bounded domain that is either convex or possesses a* $C^{1,1}$*-boundary (cf.* [10, Section 6.2]*),*
2) $j\colon V \to \mathbb{R}$ *is weakly lower semicontinuous, twice continuously differentiable and bounded from below,*
3) $\sigma > 0$,
4) $p \in \mathbb{N} \setminus \{0\}$, $\mathsf{A} \in \mathbb{R}^{p\times p}$ *symmetric, positive definite and* $\|\cdot\|_{\mathsf{A}} := \langle \mathsf{A}\cdot, \cdot\rangle^{1/2}$,
5) $\mathcal{B} : \mathbb{R}^p \to L^2(\Omega)$ *is linear and bounded.*

The governing constraint is a rather well understood semilinear elliptic PDE – cf. [8, 9], for instance. It features a "level-1-type" nonsmooth Nemytski-operator that induces a nonsmoothness in the solution operator, which remains Lipschitz continuous and Hadamard differentiable. Depending on the specific result, our examinations can, of course, be generalized for additional parameters or different Nemytski-operators. We refer to Assumption 1.1 and Section 2 for the exact setting chosen in this paper. Some examples of problems related to the nonsmooth PDE arise, for instance, in mechanics, plasma physics and the context of certain combustion processes, see, e.g., [16, 21, 22, 25] for possible areas of application.

The goal of the present paper is the development, analysis and numerical realization of efficient, fast and certified solution algorithms for the necessary stationarity system of (**P**). Of course, for the implementation, a discretization of the state space V and the PDE constraint is required. Utilizing a standard finite element (FE) method for the discretization of the stationarity system of (**P**), we derive a complex nonlinear and nonsmooth large-scale system. To obtain fast numerical solution methods, a reduced order approach is applied employing the reduced basis (RB) method – cf., e.g., [3, 13, 18, 20].

For the construction of an accurate and efficient RB scheme, a-posteriori error analysis is required. Recall that the RB method is especially efficient for parametrized linear-quadratic optimal control problems – see [15, 17], for instance. For RB results in the framework of nonlinear elliptic PDEs, we refer, e.g., to [5, 11, 23] and to the recent work [14].

Results concerning model order reduction (MOR) for nonsmooth PDE constrained optimization are rarely found in the literature, with the only contribution, to the best of the authors' knowledge, being [4], where an offline/online based (greedy) reduced basis framework for the PDE constraint in (**P**) is investigated. The results showed that the RB method combined with empirical interpolation techniques (cf., e.g., [2, 6, 7]) can improve the efficiency of solving the PDE – with nonsmooth effects being the limiting factor for the quality of the reduced order approximations, however. Since the constraint is part of the stationarity system of (**P**), we extend the results obtained in the thesis [4] and present an RB approach for the constraint that is justified by efficient a-posteriori error analysis. This way,

we obtain a significant reduction of the CPU times compared to a standard FE discretization. For the optimization problem (**P**), we consider strong first-order optimality conditions based on those derived in [9], i.e., the system

$$-\Delta\bar{y} + \max\{0, \bar{y}\} = \mathcal{B}\bar{\mu}, \tag{1a}$$

$$-\Delta\bar{p} + \mathbb{1}_{\{\bar{y}>0\}}\bar{p} = j'(\bar{y}), \tag{1b}$$

$$\mathcal{B}^*\bar{p} + \sigma\mathsf{A}\mu = 0 \tag{1c}$$

and the corresponding pseudo semismooth Newton (PSN) scheme. In (1b), the symbol $\mathbb{1}$ denotes the indicator function and in (1c), the operator $\mathcal{B}^* \in \mathcal{L}(L^2(\Omega), \mathbb{R}^p)$ is the adjoint of $\mathcal{B}$. Starting from a classical offline/online RB concept, we introduce a novel adaptive combined RB-PSN scheme that does not require an offline phase and improves the RB approximation quality simultaneously with the solution iterations using local information along the iterates.

The paper is organized as follows: In Section 2, we introduce and analyze the RB method for the parametrized nonsmooth PDE constraint of the optimization problem – including residual based error estimation. Numerical tests illustrate the drastic reduction of the CPU times compared to a piecewise linear FE scheme. Section 3 covers the offline/online and the adaptive RB approach for the stationarity system (1). We explain the adaptive RB approach in detail, derive a residual based error indicator for the stationarity system and illustrate the efficiency and accuracy of the method by several numerical experiments. Finally, we draw some conclusions in Section 4.

2 MOR for the State Equation

The aim of this section is to establish classical MOR results based on the RB approach – such as error estimators, a-priori estimates and convergence theory – for the nonsmooth PDE constraint. The crucial ingredients for this will be the monotonicity of the max-operator and the Lipschitz continuity of the solution operator to the PDE. The presented results are mainly based on [4].

Let us first fix the exact analytical framework for the FE and RB analysis of the PDE and recall some of the known results. We consider the more general parametrized boundary value problem

$$c(\mu)\langle\nabla y, \nabla\varphi\rangle_{L^2} + a(\mu)\langle\max\{0, y\}, \varphi\rangle_{L^2} = \langle f(\mu), \varphi\rangle_{V',V} \quad \forall\varphi \in W, \tag{2}$$

where μ is in a parameter set $\mathcal{P} \subset \mathbb{R}^p$, $p \in \mathbb{N}\setminus\{0\}$, and W denotes a closed subspace of V – in the FE setting, this is the finite element space, in the MOR framework,

this will be a low-dimensional subspace spanned by the RB functions. For the remainder of this section, we additionally assume the following:

Assumption 2.1. 1) $\mathcal{P} \subset \mathbb{R}^p$, $p \in \mathbb{N} \setminus \{0\}$ *is nonempty and closed,*
2) $c\colon \mathcal{P} \to \mathbb{R}$ *is L_c-Lipschitz continuous, positive and uniformly bounded away from zero,*
3) $a\colon \mathcal{P} \to \mathbb{R}$ *is L_a-Lipschitz continuous and nonnegative,*
4) $f\colon \mathcal{P} \to V'$ *is L_f-Lipschitz continuous,*
5) $\mathcal{P}$ *is compact or both a and c are constant.*

The following proposition sums up the standard existence and uniqueness result for solutions to equation (2) given the assumptions made above and provides some additional information on the Lipschitz constant.

Proposition 2.2. *Let Assumptions* 1.1 *and* 2.1 *be satisfied. For every $\mu \in \mathcal{P}$, there exists a unique solution $y \in W$ to the equation* (2)*. The induced solution operator $\mathcal{S}_W\colon \mathcal{P} \to W$ is $L_{\mathcal{S}}$-Lipschitz continuous with the constant*

$$L_{\mathcal{S}} = \frac{C_1}{C_2}\left(\frac{L_c + L_a C_1}{C_2} \sup_{\mu\in\mathcal{P}} \|f(\mu)\|_{V'} + L_f\right) \tag{3}$$

independently of W, where $C_P > 0$ denotes the Poincaré constant, $C_1 = 1 + C_P^2$ and $C_2 = \inf\{c(\mu) | \mu \in \mathcal{P}\}$.

Proof. Though the result is fairly standard, we will give a short outline of the proof for the sake of completeness. Existence and uniqueness is guaranteed by the Browder-Minty theorem. For the proof of the Lipschitz continuity, we set $y_1 \coloneqq y(\mu_1)$ and $y_2 \coloneqq y(\mu_2)$. Without loss of generality, we assume that $\|y_1 - y_2\|_V > 0$, otherwise, the statement is trivial. We subtract the PDEs

$$c(\mu_1)\langle\nabla y_1, \nabla\varphi\rangle_{L^2} + a(\mu_1)\langle\max\{0, y_1\}, \varphi\rangle_{L^2} = \langle f(\mu_1), \varphi\rangle_{V',V} \quad \forall\varphi \in W, \tag{4}$$

$$c(\mu_2)\langle\nabla y_2, \nabla\varphi\rangle_{L^2} + a(\mu_2)\langle\max\{0, y_2\}, \varphi\rangle_{L^2} = \langle f(\mu_2), \varphi\rangle_{V',V} \quad \forall\varphi \in W \tag{5}$$

add a zero and rearrange the terms to obtain

$$\begin{aligned}&\langle c(\mu_1)\nabla y_1 - c(\mu_2)\nabla y_2, \nabla\varphi\rangle_{L^2} + a(\mu_1)\langle\max\{0, y_1\} - \max\{0, y_2\}, \varphi\rangle_{L^2}\\ &+ (a(\mu_1) - a(\mu_2))\langle\max\{0, y_2\}, \varphi\rangle_{L^2} = \langle f(\mu_1) - f(\mu_2), \varphi\rangle_{V',V}.\end{aligned}$$

Now, we test with $\varphi = y_1 - y_2 \in W$, use the monotonicity of the max operator, the Cauchy-Schwarz inequality and nonnegativity of a to obtain that

$$\begin{aligned}&c(\mu_1)\|\nabla(y_1 - y_2)\|_{L^2}^2 \le |a(\mu_1) - a(\mu_2)|\|\max\{0, y_2\}\|_{L^2}\|y_1 - y_2\|_{L^2}\\ &+ \|f(\mu_1) - f(\mu_2)\|_{V'}\|y_1 - y_2\|_V + |c(\mu_1) - c(\mu_2)|\|\nabla y_2\|_{L^2}\|\nabla(y_1 - y_2)\|_{L^2}.\end{aligned}$$

The Poincaré inequality yields the V-estimate

$$\frac{c(\mu_1)}{C_1}\|y_1 - y_2\|_V^2 \leq |a(\mu_1) - a(\mu_2)|\|\max\{0, y_2\}\|_{L^2}\|y_1 - y_2\|_{L^2}$$
$$+ \|f(\mu_1) - f(\mu_2)\|_{V'}\|y_1 - y_2\|_V + |c(\mu_1) - c(\mu_2)|\|\nabla y_2\|_{L^2}\|\nabla(y_1 - y_2)\|_{L^2},$$

where the Lipschitz continuity of the coefficient functions gives

$$\frac{c(\mu_1)}{C_1}\|y_1 - y_2\|_V^2 \leq (L_a\|y_2\|_{L^2} + L_f + L_c\|\nabla y_2\|_{L^2})\|y_1 - y_2\|_V\|\mu_1 - \mu_2\|_{\mathbb{R}^p}. \tag{6}$$

Applying the Poincaré inequality once more, we obtain the estimate

$$\|y_2\|_V^2 \leq C_1\|\nabla y_2\|_{L^2}^2. \tag{7}$$

We test (5) with $\varphi = y_2$ and end up with

$$\langle c(\mu_2)\nabla y_2, \nabla y_2\rangle_{L^2} + \underbrace{\langle a(\mu_2)\max\{0, y_2\}, y_2\rangle_{L^2}}_{\geq 0} = \langle f(\mu_2), y_2\rangle_{V',V}$$

due to the monotonicity of the max operator, thus

$$c(\mu_2)\|\nabla y_2\|_{L^2}^2 \leq \langle f(\mu_2), y_2\rangle_{V',V}. \tag{8}$$

Now, we combine equations (7) and (8), obtaining that

$$\|y_2\|_V \leq \frac{C_1}{c(\mu_2)}\|f(\mu_2)\|_{V'}.$$

We combine that with (6) and (8), divide by $\|y_1 - y_2\|_V$, use positivity of the lower bound of c and rearrange terms to obtain that

$$\|y_1 - y_2\|_V \leq \frac{C_1}{C_2}\left(\frac{L_c + L_aC_1}{C_2}\|f(\mu_2)\|_{V'} + L_f\right)\|\mu_1 - \mu_2\|_{\mathbb{R}^p}.$$

If a and c are constant, then $L_a = L_c = 0$, which yields (3) with the usual convention of $0 \cdot \infty = 0$. Otherwise $\mathcal{P}$ is compact and we can form the supremum, which is attained as a maximum due to the compactness. □

Remark 2.3. 1) *The Lipschitz constant $L_{\mathcal{S}}$ is explicitly computable for a given domain Ω whose Poincaré constant C_P is known. This can be used to get the quantities in the convergence rate of the RB-to-FE distance, cf. Theorem* 2.7.

2) *Higher regularity of the right hand side f induces higher regularity of the solutions. In this section, we are not concerned with these effects. The upcoming optimization section, however, is quite reliant on the corresponding effect in the optimization formulation.*

3) *Notice that the solution operator $\mathcal{S}_W$ is generally not Gâteaux differentiable (cf.* [9, Section 6]*), but Hadamard differentiable for a and c constant, as shown in* [8, Theorem 22].

2.1 RB Analysis for the State Equation

For the RB analysis, we will assume that we are able to approximate the analytical solution of (2) with $W = V$ arbitrarily well using finite dimensional FE spaces $V_h \subset V$ endowed by the V-topology. Accordingly, the solution $y_h(\mu)$ of (2) with $W = V_h$ will be considered the reference solution. We follow the classical offline/online splitting approach (cf., e.g., [12, 18]), where the RB space is generated via a greedy algorithm in an offline phase so that the RB solutions with respect to this basis can be computed quickly later on in the online phase. The offline basis computation for an error indicator $\Delta(V_\ell, \mu)$ that satisfies

$$\|y_h(\mu) - y_\ell(\mu)\|_V \le \Delta(V_\ell, \mu) \quad \text{for all } \mu \in \mathcal{P}_{\text{train}},$$

where $y_h(\mu) \in V_h$ and $y_\ell(\mu) \in V_\ell$ denote the FE and RB solutions, respectively, is described in Algorithm 1.

Algorithm 1: Greedy RB Method for PDE.

Require : Discrete training set of parameters $\mathcal{P}_{\text{train}} \subset \mathcal{P}$,
error tolerance $\varepsilon_{tol} > 0$

Return : RB parameters $\mathcal{P}_\ell$, reduced basis Ψ_ℓ, RB space V_ℓ

Set $\ell = 0$, $\mathcal{P}_0 = \emptyset$, $\Psi_0 = \emptyset$, $V_0 = \{0\}$;

while $\varepsilon_\ell := \max\{\Delta(V_\ell, \mu) \,|\, \mu \in \mathcal{P}_{\text{train}}\} > \varepsilon_{tol}$ **do**

- Compute $\mu_{\ell+1} \in \arg\max\{\Delta(V_\ell, \mu) \,|\, \mu \in \mathcal{P}_{\text{train}}\}$;
- Set $\mathcal{P}_{\ell+1} = \mathcal{P}_\ell \cup \{\mu_{\ell+1}\}$ and $\psi_{\ell+1} = y_h(\mu_{\ell+1})$;
- Orthonormalize $\psi_{\ell+1}$ against Ψ_ℓ;
- Set $\Psi_{\ell+1} = \Psi_\ell \cup \{\psi_{\ell+1}\}$;
- Define $V_{\ell+1} = V_\ell \oplus \operatorname{span}(\psi_{\ell+1})$ and $\ell = \ell + 1$;

Note that the indicator $\Delta(V_\ell, \mu)$ is an essential component of the algorithm and should be easily evaluable. In our case, the monotonicity of the max term allows us to derive a residual-based error estimator that only depends on the RB solution and can thus be evaluated without solving the possibly costly FE discretized PDE.

Proposition 2.4. *Let Assumptions* 1.1 *and* 2.1 *hold and* $e_y(\mu) := y_h(\mu) - y_\ell(\mu)$. *Then*

$$\|e_y(\mu)\|_V \le \Delta_w^y(V_\ell, \mu) := \frac{C_1}{C_2} \|\mathrm{Res}_\ell^y(\mu)\|_{V_h'} \tag{9}$$

for every $\mu \in \mathcal{P}$ *with* $C_1 = 1 + C_P^2$, $C_2 = \inf\{c(\mu) | \mu \in \mathcal{P}\}$ *and the residual* $\mathrm{Res}_\ell^y : \mathcal{P} \to V_h'$ *given by*

$$\begin{aligned}\langle \mathrm{Res}_\ell^y(\mu), \varphi \rangle_{V_h', V_h} := \langle f(\mu), \varphi \rangle_{V',V} - c(\mu) \langle \nabla y_\ell(\mu), \nabla \varphi \rangle_{L^2} \\ - a(\mu) \langle \max\{0, y_\ell(\mu)\}, \varphi \rangle_{L^2} \quad \text{for all } \varphi \in V_h.\end{aligned}$$

Proof. Let $\mu \in \mathcal{P}$ be fixed. Then the FE solution $y_h(\mu)$ satisfies the equation

$$c(\mu)\langle \nabla y_h(\mu), \nabla \varphi \rangle_{L^2} + a(\mu)\langle \max\{0, y_h(\mu)\}, \varphi \rangle_{L^2} = \langle f(\mu), \varphi \rangle_{V',V} \tag{10}$$

for all $\varphi \in V_h$ and $e_y(\mu)$ is in V_h. We use the Poincaré inequality, the monotonicity of the max term and (10) to obtain that

$$\begin{aligned}
&\frac{C_2}{C_1}\|e^y(\mu)\|_V^2 \le c(\mu)\|\nabla e^y(\mu)\|_{L^2}^2 \\
&\le c(\mu)\|\nabla e^y(\mu)\|_{L^2}^2 + a(\mu)\langle \max\{0, y_h(\mu)\} - \max\{0, y_\ell(\mu)\}, y_h(\mu) - y_\ell(\mu)\rangle_{L^2} \\
&= c(\mu)\langle \nabla y_h(\mu), \nabla e^y(\mu)\rangle_{L^2} + a(\mu)\langle \max\{0, y_h(\mu)\}, e^y(\mu)\rangle_{L^2} \\
&\quad - c(\mu)\langle \nabla y_\ell(\mu), \nabla e^y(\mu)\rangle_{L^2} - a(\mu)\langle \max\{0, y_\ell(\mu)\}, e^y(\mu)\rangle_{L^2} \\
&= \langle f(\mu), e^y(\mu)\rangle_{V',V} - c(\mu)\langle \nabla y_\ell(\mu), \nabla e^y(\mu)\rangle_{L^2} - a(\mu)\langle \max\{0, y_\ell(\mu)\}, e^y(\mu)\rangle_{L^2} \\
&= \langle \mathrm{Res}_\ell^y(\mu), e^y(\mu)\rangle_{V_h',V_h} \le \|\mathrm{Res}_\ell^y(\mu)\|_{V_h'}\|e^y(\mu)\|_V,
\end{aligned}$$

which gives (9). □

Remark 2.5. *The error estimation requires us to compute the dual norm of the residual, which is equal to the primal norm of the V-Riesz representative of the residual. To compute the Riesz representative, we need to solve a FE discretized linear elliptic PDE. Depending on the exact application, this may end up being more costly than computing the true error, e.g., when the reduced basis of the greedy RB method ends up being large in comparison to the size of the training set $\mathcal{P}_{\mathrm{train}}$ and the semilinear PDE can be solved with few semismooth Newton iterations.*

As a direct consequence of Proposition 2.4, we obtain the reproduction-of-solutions property and the vanishing-error-bound property, which are classical RB results.

Corollary 2.6. *Let Assumptions* 1.1 *and* 2.1 *hold. Assume that for a given $\mu \in \mathcal{P}$ the FE solution $y_h(\mu) \in V_h$ belongs to V_ℓ. Then the corresponding RB solution $y_\ell(\mu) \in V_\ell$ satisfies*

1) $y_h(\mu) = y_\ell(\mu)$,
2) $\Delta_w^y(V_\ell, \mu) = 0$.

Proof. Part 1) follows from the proof of Proposition 2.4. Furthermore, we derive Part 2) directly from the definition of the residual and Part 1). □

Similarly, we can show convergence of the RB solutions to the FE solution, as the parameters used in the reduced basis fill the parameter set.

Theorem 2.7. *Let Assumptions* 1.1 *and* 2.1 *hold and let* $(\mathcal{P}_\ell)_{\ell\in\mathbb{N}} \subset \mathcal{P}$ *denote a sequence of parameter subsets and* $h_\ell := \sup\{\operatorname{dist}(\mu, \mathcal{P}_\ell) \mid \mu \in \mathcal{P}\}$. *Then*

$$\sup_{\mu\in\mathcal{P}} \|y_h(\mu) - y_\ell(\mu)\|_V \in \mathcal{O}(h_\ell).$$

Proof. For $\ell \in \mathbb{N}$ and $\mu \in \mathcal{P}$, we can choose $\mu^* \in \arg\min\{\|\mu - \tilde{\mu}\|_{\mathbb{R}^p} \mid \tilde{\mu} \in \mathcal{P}_\ell\}$. Now, we use the Lipschitz continuity of the solution operator $\mathcal{S}_W$ and its independence of the chosen space W (Proposition 2.2) and the reproduction of solutions property (Corollary 2.6 *1)*) to obtain that

$$\begin{aligned}
&\|y_h(\mu) - y_\ell(\mu)\|_V \\
&\quad \le \|y_h(\mu) - y_h(\mu^*)\|_V + \|y_h(\mu^*) - y_\ell(\mu^*)\|_V + \|y_\ell(\mu) - y_\ell(\mu^*)\|_V \\
&\quad \le 2L_S \|\mu - \mu^*\|_{\mathbb{R}^p} \le 2L_S h_\ell,
\end{aligned}$$

which yields the claim. □

Hence, if the sequence $\mathcal{P}_\ell$ gets dense in $\mathcal{P}$, i.e., $\lim_{\ell\to\infty} h_\ell = 0$, the RB solutions converge to the FE solution. As a direct consequence, we can obtain an a-priori error bound.

Corollary 2.8. *Let Assumptions* 1.1 *and* 2.1 *hold, let* $\mathcal{P}_{\text{train}} \subset \mathcal{P}$ *be a discrete set with filling distance* h_{train} *and let the RB greedy algorithm (Algorithm* 1*) terminate with tolerance* ε_{tol}. *Then*

$$\|y_h(\mu) - y_\ell(\mu)\|_V \le 2L_{\mathcal{S}} h_{train} + \varepsilon_{tol}.$$

Proof. Analogously to the proof of Theorem 2.7. □

2.2 Numerical Results for the State Equation

As we have seen in Section 2, several classical RB results hold for the PDE in question – including convergence and error estimation results. The proofs of these results were virtually independent of the nonsmoothness property of the max operator and the solution operator to the PDE. The crucial properties were the monotonicity of the max operator and the Lipschitz continuity of the solution operator to the PDE. We will confirm the results of the previous section numerically in this one and use a carefully constructed example to show that the nonsmoothness effects of the PDE are in fact the limiting factor for the RB performance in practice. Additionally, we will show a theoretical bound on the condition number of the RB system. Let us mention that the example settings introduced in this section will be reused in Section 3.4.

For our numerical experiments, we fix the domain $\Omega = (0,1)^2$ and consider P_1-type FE on a Friedrichs-Keller triangulation of the domain. The measure of fineness of the grids will be $h > 0$, which denotes the inverse number of square cells per dimension – i.e., the grid will have $2/h^2$ triangles. From here on out, we write the coefficient vector of the piecewise linear interpolant on the grid vertices of a function $z\colon \Omega \to \mathbb{R}$ in typewriter font (i.e., $\mathsf{z} \in \mathbb{R}^N$) and use the same font for the matrices in the discretized settings. Dealing with the nonlinear max term, we resort to mass lumping for this term in order to be able to evaluate it componentwisely. Inevitably, this introduces a numerical discretization error. Its effects decrease with increasing fineness of the discretization but increase with the coefficient function a that scales the nonlinearity. The corresponding stiffness matrix $\mathsf{K} \in \mathbb{R}^{N\times N}$, mass matrix $\mathsf{M} \in \mathbb{R}^{N\times N}$, lumped mass matrix $\tilde{\mathsf{M}} \in \mathbb{R}^{N\times N}$ and the right hand side $\mathsf{f}(\mu) \in \mathbb{R}^N$ are given from the FE ansatz functions φ_i, $i = 1,\dots,N$ as

$$\mathsf{K}_{i,j} = \left(\left(\int_\Omega \nabla\varphi_i \cdot \nabla\varphi_j \mathrm{d}\boldsymbol{x}\right)\right), \quad \mathsf{f}_i(\mu) = \left(\langle f(\mu), \varphi_i\rangle_{V',V},$$

$$\mathsf{M}_{i,j} = \left(\left(\int_\Omega \varphi_i\varphi_j \mathrm{d}\boldsymbol{x}\right)\right), \qquad \tilde{\mathsf{M}} = \operatorname{diag}\left(\frac{1}{3}\left|\operatorname{supp}(\varphi_i)\right| \,:\, i = 1,\dots,N\right).$$

The discretization leaves us with the following FE and RB systems:

$$\begin{aligned} c(\mu)\mathsf{K}\mathsf{y}_h(\mu) \qquad &+ a(\mu)\tilde{\mathsf{M}}\max\{0,\mathsf{y}_h(\mu)\} \qquad - \mathsf{f}(\mu) \quad = 0 \quad \text{in } \mathbb{R}^N,\\ c(\mu)\boldsymbol{\Psi}_\ell^T\mathsf{K}\boldsymbol{\Psi}_\ell\mathsf{y}_\ell(\mu) &+ a(\mu)\boldsymbol{\Psi}_\ell^T\tilde{\mathsf{M}}\max\{0,\boldsymbol{\Psi}_\ell\mathsf{y}_\ell(\mu)\} - \boldsymbol{\Psi}_\ell^T\mathsf{f}(\mu) = 0 \quad \text{in } \mathbb{R}^\ell, \end{aligned}$$

where $\boldsymbol{\Psi} \in \mathbb{R}^{N\times\ell} = [\boldsymbol{\psi}_1|\dots|\boldsymbol{\psi}_\ell]$ is the matrix whose columns are the FE coefficient vectors of the reduced basis functions. These finite dimensional systems are solved with a standard semismooth Newton method, where the iteration matrices are

$$\begin{aligned} \mathsf{H}_h(\mathsf{y}_h(\mu)) &\coloneqq c(\mu)\mathsf{K} \qquad + a(\mu)\tilde{\mathsf{M}}\Theta(\mathsf{y}_h(\mu)),\\ \mathsf{H}_\ell(\mathsf{y}_\ell(\mu)) &\coloneqq c(\mu)\boldsymbol{\Psi}_\ell^T\mathsf{K}\boldsymbol{\Psi}_\ell + a(\mu)\boldsymbol{\Psi}_\ell^T\tilde{\mathsf{M}}\Theta(\boldsymbol{\Psi}_\ell\mathsf{y}_\ell(\mu))\boldsymbol{\Psi}_\ell, \end{aligned}$$

with $\Theta \coloneqq \Theta_0$, where $\Theta_x\colon \mathbb{R}^N \to \mathbb{R}^{N\times N}$ maps a vector to the diagonal matrix that takes the Heaviside function with functional value x at 0 evaluated for each entry of the vector as its diagonal entries.

The RB Newton matrix possesses the favorable property that its condition number is bounded independently of the current iterate and the dimension of the RB system.

Proposition 2.9. *Let $\mathsf{Q} \in \mathbb{R}^{N\times N}$ be symmetric and positive definite and let $\boldsymbol{\Psi}_\ell \in \mathbb{R}^{N\times\ell}$ represent a reduced basis that is orthonormal with respect to the scalar product induced by Q. Then the condition number of the Newton matrix $\mathsf{H}_\ell(\mathsf{y}_\ell(\mu))$ is bounded independently of ℓ, y_ℓ and μ.*

Proof. First, we fix ℓ, y_ℓ and μ and define $\mathsf{H} = \mathsf{H}_\ell(\mathsf{y}_\ell(\mu)) \in \mathbb{R}^{\ell\times\ell}$. Since K is symmetric and positive definite, the matrix $\tilde{\mathsf{M}}\Theta(\Psi_\ell \mathsf{y}_\ell(\mu))$ is symmetric and positive semidefinite, the functions $c(\mu) > 0$, $a(\mu) \geq 0$ and $\operatorname{rank}(\Psi_\ell) = \ell$, the matrix H is symmetric and positive definite. Therefore, we can compute $\operatorname{cond}_2(\mathsf{H}) = \lambda_{\max}/\lambda_{\min}$, where $\lambda_{\max} \geq \lambda_{\min} > 0$ are the largest and smallest eigenvalue, respectively, of H. For an element $\mathsf{v} \in \mathbb{R}^\ell$ with

$$\mathsf{w} \coloneqq \Psi_\ell \mathsf{v} = \sum_{i=1}^{\ell} \mathsf{v}_i \boldsymbol{\psi}_i \in \mathbb{R}^N,$$

where $\boldsymbol{\psi}_i$ is the i-th column of Ψ_ℓ, we define $\|\mathsf{w}\|_{\mathsf{Q}} = \langle \mathsf{Q}\mathsf{w}, \mathsf{w}\rangle_{\mathbb{R}^N}^{1/2}$. This gives

$$\|\mathsf{w}\|_{\mathsf{Q}}^2 = \langle \mathsf{Q}\mathsf{w}, \mathsf{w}\rangle_{\mathbb{R}^N} = \sum_{i,j=1}^{\ell} \mathsf{v}_i \mathsf{v}_j \underbrace{\langle \mathsf{Q}\boldsymbol{\psi}_i, \boldsymbol{\psi}_j\rangle_{\mathbb{R}^N}}_{=\delta_{ij}} = \sum_{i=1}^{\ell} \mathsf{v}_i^2 = \|\mathsf{v}\|_{\mathbb{R}^\ell}^2.$$

Since all finite dimensional norms are equivalent, there are (generally N-dependent) constants $\kappa_1, \kappa_2 > 0$ such that

$$\frac{1}{\kappa_1}\|\cdot\|_{\mathsf{Q}} \leq \|\cdot\|_{\mathsf{K}} \leq \kappa_1 \|\cdot\|_{\mathsf{Q}}, \quad \frac{1}{\kappa_2}\|\cdot\|_{\mathsf{Q}} \leq \|\cdot\|_{\mathbb{R}^N} \leq \kappa_2 \|\cdot\|_{\mathsf{Q}}.$$

Now let $\mathsf{v} \in \mathbb{R}^\ell$ be an eigenvector of H for eigenvalue $\lambda_{\max}$. We have that

$$\begin{aligned}
\lambda_{\max}\|\mathsf{v}\|_{\mathbb{R}^\ell}^2 &= \langle \mathsf{H}\mathsf{v}, \mathsf{v}\rangle_{\mathbb{R}^\ell} \\
&= c(\mu)\langle \mathsf{K}\mathsf{w}, \mathsf{w}\rangle_{\mathbb{R}^N} + a(\mu)\langle \tilde{\mathsf{M}}\Theta(\Psi_\ell \mathsf{y}_\ell(\mu))\mathsf{w}, \mathsf{w}\rangle_{\mathbb{R}^N} \\
&\leq c(\mu)\|\mathsf{w}\|_{\mathsf{K}}^2 + a(\mu)\|\tilde{\mathsf{M}}\|\|\mathsf{w}\|_{\mathbb{R}^N}^2 \leq c(\mu)\kappa_1^2\|\mathsf{w}\|_{\mathsf{Q}}^2 + a(\mu)\|\tilde{\mathsf{M}}\|\kappa_2^2\|\mathsf{w}\|_{\mathsf{Q}}^2.
\end{aligned}$$

Since $\|\mathsf{w}\|_{\mathsf{Q}} = \|\mathsf{v}\|_{\mathbb{R}^\ell}$, we can conclude that

$$\lambda_{\max} \leq c_1 \coloneqq \kappa_1^2 \max_{\mu\in\mathcal{P}} c(\mu) + \kappa_2^2\|\tilde{\mathsf{M}}\| \max_{\mu\in\mathcal{P}} a(\mu) < \infty.$$

Now let $\mathsf{v} \in \mathbb{R}^\ell$ be an eigenvector of H corresponding to $\lambda_{\min}$. Then, since $\tilde{\mathsf{M}}\Theta(\Psi_\ell \mathsf{y}_\ell(\mu))$ is positive semidefinite, we can conclude that

$$\begin{aligned}
\lambda_{\min}\|\mathsf{v}\|_{\mathbb{R}^\ell}^2 = \langle \mathsf{H}\mathsf{v}, \mathsf{v}\rangle_{\mathbb{R}^\ell} &= c(\mu)\langle \mathsf{K}\mathsf{w}, \mathsf{w}\rangle_{\mathbb{R}^N} + \underbrace{a(\mu)\langle \tilde{\mathsf{M}}\Theta(\Psi_\ell \mathsf{y}_\ell(\mu))\mathsf{w}, \mathsf{w}\rangle_{\mathbb{R}^N}}_{\geq 0} \\
&\geq c(\mu)\langle \mathsf{K}\mathsf{w}, \mathsf{w}\rangle_{\mathbb{R}^N} \geq \frac{c(\mu)}{\kappa_1^2}\|\mathsf{w}\|_{\mathsf{Q}}^2.
\end{aligned}$$

Again, using $\|\mathsf{w}\|_{\mathsf{Q}} = \|\mathsf{v}\|_{\mathbb{R}^\ell}$, we can conclude that

$$\lambda_{\min} \geq c_2 \coloneqq \frac{\min\{c(\mu)|\mu \in \mathcal{P}\}}{\kappa_1^2} > 0,$$

which yields $\operatorname{cond}_2(\mathsf{H}) \leq c_1/c_2$ for c_1 and c_2 independent of ℓ, y_ℓ and μ. □

Our code is implemented in Python3 and uses FENICS [1] for the matrix assembly. Sparse memory management and computations are implemented with SciPy [24]. All computations below were run on an Ubuntu 18.04 notebook with 12 GB main memory and an Intel Core i7-4600U CPU.

Example 2.1. For the first numerical example, we set $\mathcal{P} = [-2, 2]^2$, $c(\mu) = 1$, $a(\mu) = 10$ and the right hand sides $f(\mu)$ (to be understood as L^2-functions mapping $(x_1, x_2) \in \Omega$ to $\mathbb{R}$ embedded into V') as

$$f(\mu)(x_1, x_2) = 10 \begin{cases} \mu_1 x_1 x_2, & \text{for } \boldsymbol{x} = (x_1, x_2) \text{ and } x_1 \leq \frac{1}{2}, \\ \mu_2 x_1^2 x_2^2, & \text{otherwise.} \end{cases} \tag{11}$$

The right hand side and corresponding solution are shown in Figure 1 for two values of μ. Note that the right hand side is linear in the parameter μ.

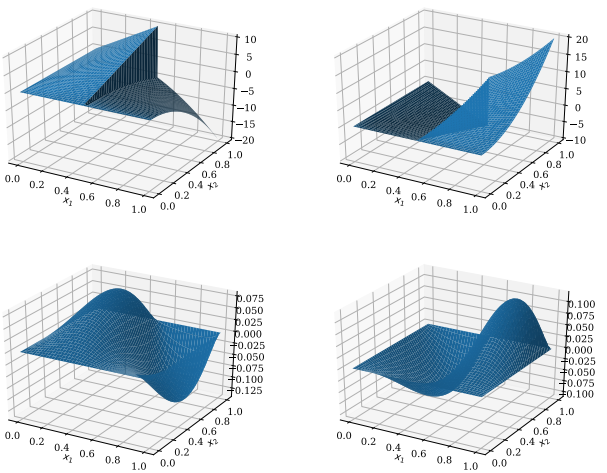

Fig. 1: Example 2.1. Right hand side (top) and FE solution (bottom) for parameters $\mu = (2, -2)$ (left column) and $\mu = (-2, 2)$ (right column) for $1/h = 400$.

We fix a training set of 121 and a test set of 196 equidistant points in $\mathcal{P}$, where both sets are chosen disjointly. We solve the FE system for each parameter in the test set and compare the results to those from solving the corresponding RB systems generated from offline phases that use the true V-error or the error estimator on the training set, respectively. The semismooth Newton iterations in both the FE and the RB computations for a parameter are started using the solution for the closest parameter in the test set whose solution is already known as an initial value

(warm starting). The semismooth Newton method terminates when the tolerance of 10^{-8} is reached for the dual norm of the residual and the RB error tolerance is set to 10^{-4}. The results for different grids with step size $h = 1/100, 1/200, 1/400$ per dimension can be found in Table 1.

FE Results		
1/h	**avg. iterations**	**avg. time (s)**
100	2.03	0.49
200	2.03	3.22
400	2.02	21.88

offline/online RB (true V-error)					
1/h	**avg. iterations**	**avg. speed-up**	t_{off} **(s)**	**avg. error**	$\lvert\Psi\rvert$
100	2.02	71.00	70.00	$2.38 \cdot 10^{-5}$	17
200	2.02	127.60	402.90	$2.34 \cdot 10^{-5}$	17
400	2.02	250.36	2793.68	$2.32 \cdot 10^{-5}$	17

offline/online RB (estimator)					
1/h	**avg. iterations**	**avg. speed-up**	t_{off} **(s)**	**avg. error**	$\lvert\Psi\rvert$
100	2.02	51.38	379.02	$1.15 \cdot 10^{-5}$	26
200	2.02	98.60	2264.12	$1.11 \cdot 10^{-5}$	26
400	2.02	189.65	16994.40	$1.10 \cdot 10^{-5}$	26

Tab. 1: Example 2.1. The first part of the table shows the average number of required semismooth Newton iterations for computing the FE solutions to the PDE for each parameters in the test set and the average required time. The second and third parts show the average number of semismooth Newton iterations, the average online speed-up, the required computational time for the offline phase t_{off}, the average true error $\|y_h - y_\ell\|_V$ and the size of the reduced basis generated in the offline phase with the true error and error estimator, respectively.

First of, we can note that all of the average V-errors on the test set are below the given tolerance of the offline phase that generated the reduced basis from the training set. Further, the average number of semismooth Newton iterations for solving the FE/RB systems appears to be independent of the step size h and very close for the FE and RB systems. Additionally, in this example, the size of the reduced basis is independent of the step size h (compare with Table 3 for different results). Keep in mind that the very small number of about two semismooth Newton steps per system solve is due to the warm starting. Without warm starting, the average number of semismooth Newton steps per solve is around four to five iterations with the zero function as an initial guess. As the number of cells per

dimension doubles in our implementation, the speed-up through RB on the test set approximately doubles as well, reaching up to a factor of 250. This means that if 130 or more online solves are required, the RB approach including its offline cost comes out ahead of the FE approach in computational time. As mentioned in Remark 2.5, due to the nonlinearity of the PDE, an evaluation of the error estimator means that we have to solve a linear PDE on FE level. As the number of semismooth Newton iterations per nonlinear solve is quite low in this example, it turns out that the RB approach based on the true V-error ends up outperforming the approach based on the error estimator in the offline phase and, as it works on a smaller reduced basis, in the online phase as well. Nonetheless, the error estimator is useful to estimate the error in the online phase as shown in Table 2. The effectivity – the quotient of the estimated and the true error – appears to be

1/h	avg. true V-error	avg. effectivity	time true V-error (s)	speed-up estimator
100	$2.38 \cdot 10^{-5}$	1.97	96.69	4.32
200	$2.34 \cdot 10^{-5}$	1.99	628.16	4.10
400	$2.32 \cdot 10^{-5}$	1.99	4242.60	4.10

Tab. 2: Example 2.1. Performance of true error and error estimator with reduced basis from strong greedy on test set.

mesh independent. The speed-up is approximately four. This independence of the step size can be expected, since the main difference between true error and error estimator is whether a linear or a nonlinear PDE needs to be solved on FE level. Again, a larger speed-up cannot be achieved, since the error estimator cannot be implemented in a parameter separable fashion. See also Remark 2.5.

Example 2.2. In the second example, we set $P = [0, 3]$, $c(\mu) = 1$ and $a(\mu) = 8\pi^2\mu$. In contrast to Example 1, the right hand side will be nonlinear in the parameter and we ensure that nonsmoothness effects occur by constructing a corresponding known solution that passes through zero on nonnegligible parts of the domain when the parameter μ varies. We construct the appropriate right hand side by applying the partial differential operator to the solution. To that end, we denote the four quarters of our domain as

$$\Omega_0 \coloneqq \left[0, \frac{1}{2}\right] \times \left[0, \frac{1}{2}\right], \qquad \Omega_1 \coloneqq \left[\frac{1}{2}, 1\right] \times \left[0, \frac{1}{2}\right],$$
$$\Omega_2 \coloneqq \left[\frac{1}{2}, 1\right] \times \left[\frac{1}{2}, 1\right], \qquad \Omega_3 \coloneqq \left[0, \frac{1}{2}\right] \times \left[\frac{1}{2}, 1\right],$$

and define

$$
\begin{aligned}
&g\colon \mathbb{R}^2 \to \mathbb{R},\ \boldsymbol{x} = (x_1, x_2) \mapsto \sin^2(2\pi x_1)\sin^2(2\pi x_2),\\
&b_1\colon \mathcal{P} \to \mathbb{R},\ \mu \mapsto 1 - 2\left(\min\left\{\max\left\{\mu, 2\right\}, 3\right\} - 2\right),\\
&b_2\colon \mathcal{P} \to \mathbb{R},\ \mu \mapsto -1 + 2\left(\min\left\{\max\left\{\mu, 1\right\}, 2\right\} - 1\right),\\
&b_3\colon \mathcal{P} \to \mathbb{R},\ \mu \mapsto 1 - 2\min\left\{\max\left\{\mu, 0\right\}, 1\right\}.
\end{aligned}
$$

We construct the sought out analytical solution of the PDE as

$$
y(\mu)\colon \Omega \to \mathbb{R}, \quad \boldsymbol{x} = (x_1, x_2) \mapsto \begin{cases} 0, & \boldsymbol{x} \in \mathcal{Q}_0,\\ b_1(\mu)g(\boldsymbol{x}), & \boldsymbol{x} \in \mathcal{Q}_1,\\ b_2(\mu)g(\boldsymbol{x}), & \boldsymbol{x} \in \mathcal{Q}_2,\\ b_3(\mu)g(\boldsymbol{x}), & \boldsymbol{x} \in \mathcal{Q}_3. \end{cases} \tag{12}
$$

The solution is designed as nonconstant in three quarters of the domain. We divide the parameter space into three equal parts with $\mathcal{P} = [0,3] = [0,1] \cup [1,2] \cup [2,3]$. As the parameter passes through either of the thirds of the parameter space, it flips the sign of a "bump" function with support in one of the quarters, which is prescribed by g. The change in the sign is given by the functions b_i, which are piecewise linear. Thus the amplitude of the bump changes linearly as well. Accordingly, the solution passes through 0 on an entire quarter of the domain whenever the parameter hits the middle of either of the thirds of the parameter interval. The solution is twice continuously differentiable in space, therefore the corresponding right hand side is given by $f(\mu) \coloneqq -\Delta y(\mu) + 8\pi^2 \mu \max\{0, y(\mu)\}$, where the Laplacian can be taken in the strong sense. The solution for some parameters and the functions b_i can be seen in Figure 2. We select a training set of 60 and a disjoint test set of 100 equidistant

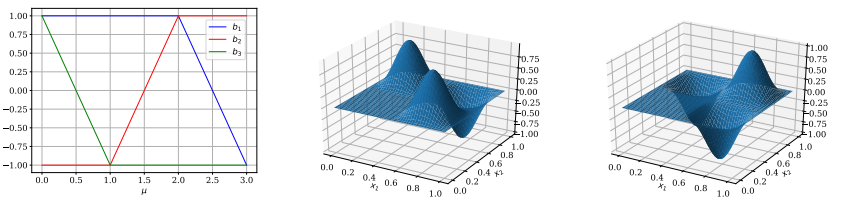

Fig. 2: Example 2.2. Functions b_i (left) and FE solution for parameters $\mu = 0$ (middle) and $\mu = 3$ (right) for $1/h = 400$.

points in $\mathcal{P} = [0,3]$ for this example. The semismooth Newton iterations in both the FE and the RB computations for a parameter are warm started as previously.

The semismooth Newton method terminates when the tolerance of 10^{-8} is reached for the dual norm of the residual and the RB error tolerance is set to 10^{-4}. The results for different grids with step size $h = 1/100, 1/200, 1/400$ per dimension can be found in Table 3. As in Example 2.1, the average RB error on the test set is

FE Results					
1/h	**avg. iterations**	**avg. time (s)**			
100	2.45	0.52			
200	2.41	3.51			
400	2.31	24.02			
offline/online RB (true V-error)					
1/h	**avg. iterations**	**avg. speed-up**	**t_{off} (s)**	**avg. error**	**$\lvert\Psi\rvert$**
100	2.40	57.54	43.57	$4.81 \cdot 10^{-5}$	24
200	2.32	131.92	266.68	$3.98 \cdot 10^{-5}$	19
400	2.26	378.26	1763.79	$2.80 \cdot 10^{-5}$	16
offline/online RB (estimator)					
1/h	**avg. iterations**	**avg. speed-up**	**t_{off} (s)**	**avg. error**	**$\lvert\Psi\rvert$**
100	2.39	47.26	177.29	$3.64 \cdot 10^{-5}$	29
200	2.37	115.46	1050.11	$2.08 \cdot 10^{-5}$	24
400	2.28	220.90	6427.29	$1.46 \cdot 10^{-5}$	19

Tab. 3: Example 2.2. The first part of the table shows the average number of required semismooth Newton iterations for computing the FE solutions to the PDE for each parameters in the test set and the average required time. The second and third parts show the average number of semismooth Newton iterations, the average online speed-up, the required computational time for the offline phase t_{off}, the average error $\|y_h - y_\ell\|_V$ and the size of the reduced basis generated in the offline phase with the true error and error estimator, respectively.

below the tolerance imposed on the training set. Additionally, the average number of semismooth Newton iterations appears to be mesh independent in this example as well. Again, the low number of average iterations is due to warm starting. As the number of cells in the grid increases, the number of functions in the reduced basis decreases in this example. This effect is unsurprising since the discretization captures the effects of the continuous PDE more accurately with increased fineness of the grids. As we know the solution, we expect three independent parts, i.e., an RB with at least three functions. When the discretization is taken to have 1024 square cells per dimension, the RB computed using the true error actually reaches this "lower bound". We do not include the full computational results for

this discretization here because the computation was run on a different system, so there is no comparability. The speed-up can be observed to roughly double as the number of cells per spatial dimension doubles as well, similarly to Example 2.1. Also as in Example 2.1, we can observe that the reduced basis approach with the true error outperforms the approach using the error estimator. The favorable effectivity of the error estimator on the test set carries over from Example 2.1, see Table 4, and the speed-up factor seems to be slightly below four as well, though the factor is not quite as stable for coarser grids in comparison to Example 2.1. This might be due to the nonconstant size of the reduced basis. Recall that the

1/h	avg. true V-error	avg. effectivity	time true V-error (s)	speed-up estimator
100	$4.81 \cdot 19^{-5}$	2.08	66.93	2.66
200	$3.98 \cdot 10^{-5}$	2.08	384.70	3.43
400	$2.80 \cdot 10^{-5}$	2.13	2489.67	3.78

Tab. 4: Example 2.2. Performance of true error and error estimator with reduced basis from strong greedy on test set.

constructed solution $y(\mu)$ to the PDE in this example has the three bumps that are flipped with the change of sign occurring at 0.5, 1.5 and 2.5 – the midpoints of the intervals $[0, 1]$, $[1, 2]$ and $[2, 3]$ that the parameter space splits up into. These are the points, where the nonlinearity and nondifferentiability of the max operator really influences the behavior of the solutions. Looking at Figure 3, we can observe that the true error and the error estimator over the test set are highest at these points of nondifferentiability, which suggests that reducing the information of the nondifferentiability in the PDE is the dominating task in the model order reduction.

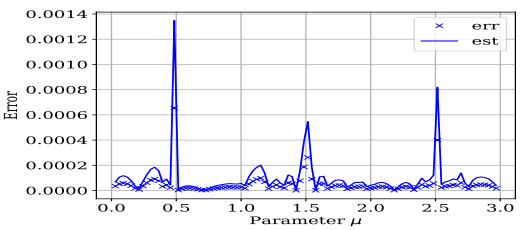

Fig. 3: Example 2.2. True error and error estimator over test set $\mathcal{P}_{\text{test}}$ with reduced basis from strong greedy on test set for $1/h = 400$.

3 MOR for the Optimal Control Problem

This section addresses the RB-based model order reduction for the PDE constrained optimal control problem (**P**) – or rather its first order optimality system. As shown in [9], given Assumption 1.1, a local minimizer $(\bar{y}, \bar{\mu}) \in V \times \mathbb{R}^p$ of the optimal control problem (**P**) satisfies the strong first order optimality system

$$\Phi(\bar{y}, \bar{p}, \bar{\mu}) \coloneqq \begin{pmatrix} -\Delta \bar{y} + \max\{0, \bar{y}\} - \mathcal{B}\bar{\mu} \\ -\Delta \bar{p} + \mathbb{1}_{\{\bar{y}>0\}}\bar{p} - j'(\bar{y}) \\ \mathcal{B}^*\bar{p} + \sigma \mathsf{A}\mu \end{pmatrix} = 0 \tag{13}$$

with an adjoint state $\bar{p} \in V$, where the system is to be understood as a problem in $V' \times V' \times \mathbb{R}^p$. By [9, Lem. 4.8], there exists at least one solution of the system given Assumption 1.1. In order to compute stationary points that satisfy this system, we follow the approach in [9] of applying a pseudo semismooth Newton (PSN) iteration that ignores the nondifferentiable indicator function in the linearization of the residual. The system matrix for this system at an iterate (y, p, μ) therefore reads as

$$\begin{pmatrix} -\Delta + \mathbb{1}_{\{y>0\}} & 0 & -\mathcal{B} \\ -j''(y) & -\Delta + \mathbb{1}_{\{y>0\}} & 0 \\ 0 & \mathcal{B}^* & \sigma \mathsf{A} \end{pmatrix}.$$

In the following, we will combine this solution procedure with reduced basis techniques in the state and the adjoint state for numerical efficiency. We derive an error estimator and -indicator for the primal and the adjoint state. Further, we present a classical offline/online RB approach and a novel adaptive RB approach that combines the PSN iterations and online updates of the reduced basis. This section should be understood as a numerical study comparing both approaches and analytical results will be limited to the error estimator and indicator.

Note that contrary to the assumptions made in the previous section, we do not assume any constraints on the parameter set $\mathcal{P} = \mathbb{R}^p$ in this section, the reason being that while we can identify a parameter region of interest in the forward problem beforehand, it is quite unclear, where the optimal parameter μ in the optimization problem will be located. However, since $\mathcal{P}$ is nonempty and closed, the setting of the stationarity system's state equation satisfies Assumption 2.1 and is therefore covered by the analysis in the previous section.

Remark 3.1. *In* [9]*, system* (13) *is reduced to the state and the adjoint state by eliminating the parameter using the last line in the system. However, the reduced system involves the operator $\mathcal{B}\mathcal{B}^*$, whose discretized matrix form is dense, introducing memory issues for our algorithms – which use sparse matrix structure when*

possible – in practice. Hence, we will keep the original three-line-system (13) *in the state, adjoint and parameter as is.*

3.1 Offline / Online RB Approach

The classical offline/online RB approach is similar to that in Section 2. As mentioned above, we are dealing with $\mathcal{P} = \mathbb{R}^p$ here, but the offline phase of the standard offline/online RB approach requires a discrete, compact training set to generate the reduced basis. In practice, we will therefore use a discrete subset of a heuristically fixed box centered at the origin that is "sufficiently large" in some sense, i.e. by including all PSN iterates as well as the optimal parameter, as the training set $\mathcal{P}_{\text{train}} \subset \mathbb{R}^p$.

Remark 3.2. *When the state dependent part of the cost functional j is bounded from below by ξ, it is possible to compute a compact set of parameters that includes all possible optimal parameters. Starting with an arbitrary $\mu_0 \in \mathbb{R}^p$, we have that*

$$\mathcal{J}(y(\mu), \mu) \geq \xi + \frac{\sigma}{2}\|\mu\|_{\mathsf{A}}^2 > \mathcal{J}(y(\mu_0), \mu_0).$$

for all $\mu \in \mathbb{R}^p$ with $\|\mu\|_{\mathsf{A}}^2 > 2(\mathcal{J}(y(\mu_0), \mu_0) - \xi)/\sigma$. Thus we can choose the set $\{\mu \in \mathbb{R}^p : \|\mu\|_{\mathsf{A}}^2 \leq 2(\mathcal{J}(y(\mu_0), \mu_0) - \xi)/\sigma\}$.

We use a common reduced basis Ψ_ℓ and reduced space V_ℓ for state and adjoint. Accordingly, we end up with the following algorithm for the offline phase of the reduced basis method in the optimal control problem.

Algorithm 2: Greedy RB Method for Stationarity System.

Require : Discrete training set of parameters $\mathcal{P}_{\text{train}} \subset \mathcal{P}$,
error tolerance $\varepsilon_{tol} > 0$

Return : RB parameters $\mathcal{P}_\ell$, reduced basis Ψ_ℓ, RB space V_ℓ

Set $\ell = 0$, $\mathcal{P}_0 = \emptyset$, $\Psi_0 = \emptyset$, $V_0 = \{0\}$;

while $\varepsilon_\ell := \max\{\Delta(V_\ell, \mu) \,|\, \mu \in \mathcal{P}_{\text{train}}\} > \varepsilon_{tol}$ **do**

- Compute $\mu_{\ell+1} \in \arg\max\{\Delta(V_\ell, \mu) \,|\, \mu \in \mathcal{P}_{\text{train}}\}$;
- Set $\mathcal{P}_{\ell+1} = \mathcal{P}_\ell \cup \{\mu_{\ell+1}\}$, $\psi_{\ell+1}^y = y_h(\mu_{\ell+1})$ and $\psi_{\ell+1}^p = p_h(\mu_{\ell+1})$;
- Orthonormalize $\left(\psi_{\ell+1}^y, \psi_{\ell+1}^p\right)$ against Ψ_ℓ;
- Set $\Psi_{\ell+1} = \Psi_\ell \cup (\{\psi_{\ell+1}^y, \psi_{\ell+1}^p\} \setminus \{0\})$;
- Define $V_{\ell+1} = V_\ell \oplus \operatorname{span}(\psi_{\ell+1}^y, \psi_{\ell+1}^p)$ and $\ell = \ell + 1$;

Again, the symbol $\Delta(V_\ell, \mu)$ denotes either the true error or an error indicator with respect to a given RB space V_ℓ. Since we apply the basis reduction on the state

and the adjoint, the true error combines both errors, i.e.,

$$e_\ell(\mu) = \begin{pmatrix} e_y(\mu) \\ e_p(\mu) \end{pmatrix}_\ell := \begin{pmatrix} y_h(\mu) - y_\ell(\mu) \\ p_h(\mu) - p_\ell(\mu) \end{pmatrix},$$
$$\|e_\ell(\mu)\| := \|y_h(\mu) - y_\ell(\mu)\|_V + \|p_h(\mu) - p_\ell(\mu)\|_V.$$

A residual based error indicator will be addressed in the section after next. Once the offline phase is completed, we solve the FE/RB discretized version of the stationarity system (13) with the PSN approach.

3.2 Adaptive RB Approach

As any of the classical offline/online greedy RB approaches, the method described in the previous subsection suffers from a number of drawbacks. The curse of dimensionality leads to an exponential growth in the offline phase's computation time, which is detrimental for high dimensional parameter spaces. This means that even if the online speed-up is significant, the computational effort of the offline phase can make the overall computational time of the RB approach exceed that of the standard FE problem. Additionally, choosing a compact trainings set that includes the optimal parameter for the offline phase is nontrivial. These drawbacks are avoided in the adaptive RB approach that we propose in this section. The method starts out with an initial starting guess for the solution and alternates PSN iterations and RB updates in order to improve the quality of the approximations of FE elements by RB elements and of the FE stationary point with the PSN iterates using local information, where local refers to the information at each of the iterates the PSN algorithm reaches. Since all of the model order reduction work is taken care of along the way, it is clear that this approach is geared towards accelerating a single solve of the stationarity system without the added expense of an offline phase – it is quite complementary to the offline/online approach in that sense. When only few, but more that one, solves of the stationarity system are required, e.g. for varying Tikhonov parameter σ in the cost functional, reusing the basis generated in the first solve is of course an option that will likely outperform offline/online approaches due to their large offline cost. The adaptive approach is somewhat similar to that in [19], where trust region methods are combined with an adaptive RB approach. Our complete algorithm is described in Algorithm 3.

The algorithm exploits a balancing of the fact that the PSN residual needs to be small only if the RB approximation is sufficiently accurate and vice versa. Starting from the initial guess, the algorithm repeatedly checks whether the RB or the PSN

Algorithm 3: Adaptive RB Method for Stationarity System.

Require : Initial parameter $\mu_0 \in \mathbb{R}^p$, error tolerances $\varepsilon_N, \varepsilon_{RB} > 0$ for RB and PSN, weight $1 > \eta > 0$, $n_{\text{fix}} \geq 0$

Return : RB approximation (y_n, p_n, μ_n) of FE stationary point (y_h, p_h, μ_h)

Set $\ell = 0$ (number of basis updates);
Set $n = 0$ (number of PSN iterations);
Set $\mathcal{P}_0 = \emptyset$, $\Psi_0 = \emptyset$, $V_0 = \emptyset$;
Set $y_0 = y_h(\mu_0)$, $p_0 = p_h(\mu_0)$;
while $\Delta(V_\ell, \mu_n) > \varepsilon_{RB}$ *or* $\|\Phi(y_n, p_n, \mu_n)\| > \varepsilon_N$ **do**
  if $\ell > 0$ *and* $\frac{(\Delta(V_\ell,\mu_n)-\varepsilon_{RB})_+}{\varepsilon_{RB}} < \eta^n \frac{(\|\Phi(y_n,p_n,\mu_n)\|-\varepsilon_N)_+}{\varepsilon_N}$ **then**
    Get $(y_{n+1}, p_{n+1}, \mu_{n+1})$ from PSN step w.r.t. Ψ_ℓ;
    Set $n = n + 1$;
  else
    Compute $\psi^y_{\ell+1} = y(\mu_n)$ and $\psi^p_{\ell+1} = p(\mu_n)$;
    Orthonormalize $(\psi^y_{\ell+1}, \psi^p_{\ell+1})$ against Ψ_ℓ;
    Set $\mathcal{P}_{\ell+1} = \mathcal{P}_\ell \cup \{\mu_n\}$, $\Psi_{\ell+1} = \Psi_\ell \cup \left(\{\psi^y_{\ell+1}, \psi^p_{\ell+1}\} \setminus \{0\}\right)$,
     $V_{\ell+1} = V_\ell \oplus \operatorname{span}\left(\psi^y_{\ell+1}, \psi^p_{\ell+1}\right)$;
    Set $\ell = \ell + 1$;
    Get $(y_{n+n_{\text{fix}}}, p_{n+n_{\text{fix}}}, \mu_{n+n_{\text{fix}}})$ from n_{fix} PSN steps w.r.t. Ψ_ℓ;
    Set $n = n + n_{\text{fix}}$;

tolerance are violated. If so, the relative violation of the tolerances is compared and depending on their magnitude, either further PSN iterations are applied or the basis is updated. We include a biasing factor η^n with $0 < \eta < 1$ in the comparison of the violations, which ensures that an improvement of the RB approximation quality is favored over additional PSN steps as the number of PSN iterations increases. In the early stages of the PSN iterations – where the residual is large and we are comparatively far from the stationary point and PSN updates are relatively coarse – the algorithm therefore benefits from small system sizes due to the small reduced bases. With increasing proximity to a stationary point, the RB approximation quality is increased to avoid pointless PSN steps on systems that do not capture the dynamics of the stationary point sufficiently well. By updating the reduced basis locally – i.e., along the PSN iterates, which are expected to exhibit behavior similar to that of the stationary point – we increase the quality of the

RB approximation of the stationary point compared to a "one-size-fits-all" type offline/online approach.

Since the evaluation of the error indicator and basis updates are computationally expensive themselves and increase the computational cost in each of the system solves of the PSN steps, we perform a fixed number n_{fix} of PSN steps whenever the reduced basis is updated to ensure that some progress towards the solution is made before the basis is expanded. This is of course an additional parameter that requires dialing in, but the numerical results in Section 3.4 suggest that choosing this parameter greater than one can indeed be beneficial. Note that, while enlarging the reduced basis increases the system size, both the basis updates and the evaluation of the true error usually gain efficiency as the iterations progress because current RB iterates (embedded into the FE space) can be used as initial guesses for the solves that have to be performed on FE level. In practice, we will see that use of the true V-error can outperform an indicator derived below, which is in part due to this behavior.

3.3 Error Estimator and Indicator

Error estimators for the RB approximation of the stationarity system need to address the errors in the state and the adjoint, i.e., $\|y_h(\mu) - y_\ell(\mu)\|_V$ and $\|p_h(\mu) - p_\ell(\mu)\|_V$. Of course, the more general results for the state in Proposition 2.4 that we derived in Section 2 carry over straightforwardly in this setting. For the adjoint state, the situation is more complicated. Due to the discontinuous indicator function in the adjoint equation, straightforward estimation is quite limited. The primary result based on Lipschitz continuity in the cost functional is summarized in the following lemma.

Lemma 3.3. *Let Assumption* 1.1 *hold and let j' be Lipschitz continuous with constant $\gamma > 0$. Then*

$$\|p_h(\mu) - p_\ell(\mu)\|_V \leq C_1 \left(\gamma \Delta_w^y(V_\ell, \mu) + \|\mathrm{Res}_\ell^p(\mu)\|_{V_h'} + \|p_\ell(\mu)\|_{L^2} \right)$$

for all $\mu \in \mathcal{P}$, where C_P is the Poincaré constant of the domain, $C_1 = 1 + C_P^2$ and the adjoint residual $\mathrm{Res}_\ell^p \colon \mathcal{P} \to V_h'$ is defined as

$$\langle \mathrm{Res}_\ell^p(\mu), \varphi \rangle_{V_h', V_h} \coloneqq \langle j'(y_\ell(\mu)), \varphi \rangle_{V',V} - \langle \nabla p_\ell(\mu), \nabla \varphi \rangle_{L^2} - \langle \mathbb{1}_{\{y_\ell(\mu) > 0\}} p_\ell(\mu), \varphi \rangle_{L^2}$$

for all $\varphi \in V_h$.

Proof. With $e_p(\mu) = p_h(\mu) - p_\ell(\mu)$, the Poincaré inequality implies that

$$\begin{aligned}\frac{\|e_p(\mu)\|_V^2}{C_1} &\leq \|\nabla\left(e_p(\mu)\right)\|_{L^2}^2 \\ &\leq \|\nabla e_p(\mu)\|_{L^2}^2 + \langle \mathbb{1}_{\{y_h(\mu)>0\}} e_p(\mu), e_p(\mu)\rangle_{L^2} \\ &= \langle j'(y_h(\mu)), e_p(\mu)\rangle_{V',V} - \langle \nabla p_\ell(\mu), \nabla e_p(\mu)\rangle_{L^2} \\ &\quad - \langle \mathbb{1}_{\{y_h(\mu)>0\}} p_\ell(\mu), e_p(\mu)\rangle_{L^2}.\end{aligned}$$

Here, we have used the adjoint equation in the last equality. Now, we utilize the assumption on j' and Proposition 2.4 to conclude that

$$\begin{aligned}\frac{\|e_p(\mu)\|_V^2}{C_1} &\leq \langle j'(y_h(\mu)) - j'(y_\ell(\mu)), e_p(\mu)\rangle_{V',V} + \langle j'(y_\ell(\mu)), e_p(\mu)\rangle_{V',V} \\ &\quad - \langle \nabla p_\ell(\mu), \nabla e_p(\mu)\rangle_{L^2} - \langle \mathbb{1}_{\{y(\mu)>0\}} p_\ell(\mu), e_p(\mu)\rangle_{L^2} \\ &\leq \gamma \|e_y(\mu)\|_V \|e_p(\mu)\|_V + \langle j'(y_\ell(\mu)), e_p(\mu)\rangle_{V',V} \\ &\quad - \langle \nabla p_\ell(\mu), \nabla e_p(\mu)\rangle_{L^2} - \langle \mathbb{1}_{\{y_\ell(\mu)>0\}} p_\ell(\mu), e_p(\mu)\rangle_{L^2} \\ &\quad + \langle \left(\mathbb{1}_{\{y_\ell(\mu)>0\}} - \mathbb{1}_{\{y_h(\mu)>0\}}\right) p_\ell(\mu), e_p(\mu)\rangle_{L^2} \\ &\leq \Big(\gamma C \|\mathrm{Res}_\ell^y(\mu)\|_{V_h'} + \|\mathrm{Res}_\ell^p(\mu)\|_{V_h'} \\ &\quad + \|(\mathbb{1}_{\{y_\ell(\mu)>0\}} - \mathbb{1}_{\{y_h(\mu)>0\}}) p_\ell(\mu)\|_{L^2}\Big) \|e_p(\mu)\|_V.\end{aligned}$$

This especially implies that

$$\begin{aligned}\|p_h(\mu) - p_\ell(\mu)\|_V \leq C_1 \Big(&\gamma \Delta_w^y(V_\ell, \mu) + \|\mathrm{Res}_\ell^p(\mu)\|_{V_h'} \\ &+ \|(\mathbb{1}_{\{y_\ell(\mu)>0\}} - \mathbb{1}_{\{y_h(\mu)>0\}}) p_\ell(\mu)\|_{L^2}\Big),\end{aligned} \tag{14}$$

and boundedness of the absolute value of the difference of the indicator functions yields the claim. □

The obvious problem with this bound is the dependence on the L_2-error of the RB adjoint state, which can cause overestimation of the true adjoint error even if we have convergence of the RB to the FE solution. The reason that we get stuck with this undesirable term is that we estimate the difference of the indicator functions in (14) very coarsely by its L^∞-upper bound. The difference of the indicator functions is intrinsically difficult to handle analytically because of their discontinuous behavior and the fact that the difference multiplied with an FE function generally is no longer in the FE space. Note, however, that using the previous result numerically, we would depend on quadrature rules in order to actually compute the L_2-norm term – in our case this would be realized by the mass lumping approach. It turns out that we can improve on the results presented in the previous lemma when

we work in this mass lumped setting. To that end, we introduce some additional notation. Since the notation for functions and their coefficient vectors are inevitably going to mix in the following, recall that we introduced the typewriter font for coefficient vectors in Subsection 2.2. For the product of an indicator function and a V_h-function z, we let $\varepsilon_{\tilde{\mathsf{M}}}(z) \coloneqq \left| \|z\|_{L_2} - (\mathsf{z}^T \tilde{\mathsf{M}} \mathsf{z})^{\frac{1}{2}} \right|$ denote the error introduced by replacing the L_2-norm with the mass lumping quadrature. Note that this is a grid dependent quantity, of course. Additionally, we introduce the shorthand notation

$$\mathsf{g}^{\pm}_{\varepsilon_{\tilde{\mathsf{M}}}}(\mu) \coloneqq \mathsf{\Psi}_\ell \mathsf{y}_\ell(\mu) \pm \left(\Delta^y_w(V_\ell, \mu) + \varepsilon_{\tilde{\mathsf{M}}}\left(y_h(\mu) - y_\ell(\mu)\right)\right) \tilde{\mathsf{M}}^{-\frac{1}{2}} \mathbf{1} \in \mathbb{R}^N, \tag{15}$$

$$\mathsf{g}^{\pm}(\mu) \coloneqq \mathsf{\Psi}_\ell \mathsf{y}_\ell(\mu) \pm \left(\Delta^y_w(V_\ell, \mu)\right) \tilde{\mathsf{M}}^{-\frac{1}{2}} \mathbf{1} \in \mathbb{R}^N, \tag{16}$$

where $\mathbf{1} \in \mathbb{R}^N$ denotes the vector of all ones. For improved readability, we will notationally suppress the functions' and vectors' dependencies on the parameter μ for the remainder of this subsection. With these quantities, we can obtain the following estimate that is based on the discrete implementation. Note that the only remaining dependence on any FE function in the right hand side of this statement is in the mass lumping error. We will disregard this dependence further down the line to obtain a reasonable (mesh dependent) error indicator.

Lemma 3.4. *Let Assumption* 1.1 *hold, j' be Lipschitz continuous with constant $\gamma > 0$ and $C_1 = 1 + C_P^2$, where C_P is the Poincaré constant. Then*

$$\|p_h(\mu) - p_\ell(\mu)\|_V \leq \Delta^p_{w,\varepsilon_{\tilde{\mathsf{M}}}}(V_\ell, \mu) \quad \textit{for every } \mu \in \mathcal{P}$$

with the term

$$\begin{aligned}
\Delta^p_{w,\varepsilon_{\tilde{\mathsf{M}}}}(V_\ell, \mu) \coloneqq C_1 \Big(& \gamma \Delta^y_w(V_\ell, \mu) + \|\mathrm{Res}^p_\ell(\mu)\|_{V_h'} + \varepsilon_{\tilde{\mathsf{M}}}\left((\mathbb{1}_{\{y_\ell > 0\}} - \mathbb{1}_{\{y_h > 0\}}) p_\ell\right) \\
& + \Big[\left(\Theta(\mathsf{g}^+_{\varepsilon_{\tilde{\mathsf{M}}}}) \mathsf{\Psi}_\ell \mathsf{p}_\ell\right)^T \tilde{\mathsf{M}} \Theta_1(-\mathsf{\Psi}_\ell \mathsf{y}_\ell) \left(\Theta(\mathsf{g}^+_{\varepsilon_{\tilde{\mathsf{M}}}}) \mathsf{\Psi}_\ell \mathsf{p}_\ell\right) \\
& + \left(\left(\Theta(\mathsf{\Psi}_\ell \mathsf{y}_\ell) - \Theta(\mathsf{g}^-_{\varepsilon_{\tilde{\mathsf{M}}}})\right) \mathsf{\Psi}_\ell \mathsf{p}_\ell \right)^T \tilde{\mathsf{M}} \Theta(\mathsf{\Psi}_\ell \mathsf{y}_\ell) \left(\left(\Theta(\mathsf{\Psi}_\ell \mathsf{y}_\ell) - \Theta(\mathsf{g}^-_{\varepsilon_{\tilde{\mathsf{M}}}})\right) \mathsf{\Psi}_\ell \mathsf{p}_\ell \right) \Big]^{\frac{1}{2}} \Big)
\end{aligned}$$

and $\Theta = \Theta_0$ with Θ_x as defined in Section 2.2.

Proof. From (14), replacing the exact L_2-norm by the mass lumped computation, we immediately obtain that

$$\begin{aligned}
\|p_h(\mu) - p_\ell(\mu)\|_V \leq C_1 \Big(& \gamma \Delta^y_w(V_\ell, \mu) + \|\mathrm{Res}^p_\ell(\mu)\|_{V_h'} \\
& + \Big[\left((\Theta(\mathsf{\Psi}_\ell \mathsf{y}_\ell) - \Theta(\mathsf{y}_h)) \mathsf{\Psi}_\ell \mathsf{p}_\ell\right)^T \tilde{\mathsf{M}} \left((\Theta(\mathsf{\Psi}_\ell \mathsf{y}_\ell) - \Theta(\mathsf{y}_h)) \mathsf{\Psi}_\ell \mathsf{p}_\ell\right) \Big]^{\frac{1}{2}} \\
& + \varepsilon_{\tilde{\mathsf{M}}}\left((\mathbb{1}_{\{y_\ell > 0\}} - \mathbb{1}_{\{y_h > 0\}}) p_\ell\right) \Big).
\end{aligned}$$

We split up the mass lumped norm term in the middle into the parts corresponding to the index sets of the nonpositive and positive nodal values of the RB state embedded into the FE space, which gives

$$\begin{aligned}
&\big((\Theta(\Psi_\ell \mathsf{y}_\ell) - \Theta(\mathsf{y}_h))\Psi_\ell \mathsf{p}_\ell\big)^T \tilde{\mathsf{M}}\big((\Theta(\Psi_\ell \mathsf{y}_\ell) - \Theta(\mathsf{y}_h))\Psi_\ell \mathsf{p}_\ell\big) \\
&= \big(\Theta(\mathsf{y}_h)\Psi_\ell \mathsf{p}_\ell\big)^T \tilde{\mathsf{M}}\Theta_1(-\Psi_\ell \mathsf{y}_\ell)\big(\Theta(\mathsf{y}_h)\Psi_\ell \mathsf{p}_\ell\big) \\
&\quad + \big((\Theta(\Psi_\ell \mathsf{y}_\ell) - \Theta(\mathsf{y}_h))\Psi_\ell \mathsf{p}_\ell\big)^T \tilde{\mathsf{M}}(\Theta(\Psi_\ell \mathsf{y}_\ell))\big((\Theta(\Psi_\ell \mathsf{y}_\ell) - \Theta(\mathsf{y}_h))\Psi_\ell \mathsf{p}_\ell\big).
\end{aligned}$$

To get rid of the dependency on the FE solution y_h, we derive componentwise upper and lower bounds that depend on RB quantities only. Using the quadrature error and the state's error estimate in Proposition 2.4, we obtain that

$$\begin{aligned}
\big[\,(\mathsf{y}_h - \Psi_\ell \mathsf{y}_\ell)^T\,\tilde{\mathsf{M}}\,(\mathsf{y}_h - \Psi_\ell \mathsf{y}_\ell)\,\big]^{\frac{1}{2}} &\le \|\,(y_h - y_\ell)\,\|_V + \varepsilon_{\tilde{\mathsf{M}}}\,(y_h - y_\ell) \\
&\le \Delta_w^y(V_\ell, \mu) + \varepsilon_{\tilde{\mathsf{M}}}\,(y_h - y_\ell) \\
&= C_1 \|\mathrm{Res}_\ell^y(\mu)\|_{V_h'} + \varepsilon_{\tilde{\mathsf{M}}}\,(y_h - y_\ell)\,,
\end{aligned}$$

and since $\tilde{\mathsf{M}}$ is a positive definite diagonal matrix, we gather that

$$|(\mathsf{y}_h - \Psi_\ell \mathsf{y}_\ell)_i| \le \frac{C_1 \|\mathrm{Res}_\ell^y(\mu)\|_{V_h'} + \varepsilon_{\tilde{\mathsf{M}}}\,(y_h - y_\ell)}{\tilde{\mathsf{M}}_{ii}^{1/2}}.$$

Accordingly, we obtain that

$$\begin{aligned}
(\mathsf{y}_h)_i &\le (\Psi_\ell \mathsf{y}_\ell)_i + |(\mathsf{y}_h - \Psi_\ell \mathsf{y}_\ell)_i| \le (\Psi_\ell \mathsf{y}_\ell)_i + \frac{C_1 \|\mathrm{Res}_\ell^y(\mu)\|_{V_h'} + \varepsilon_{\tilde{\mathsf{M}}}\,(y_h - y_\ell)}{\tilde{\mathsf{M}}_{ii}^{1/2}}, \\
(\mathsf{y}_h)_i &\ge (\Psi_\ell \mathsf{y}_\ell)_i - |(\mathsf{y}_h - \Psi_\ell \mathsf{y}_\ell)_i| \ge (\Psi_\ell \mathsf{y}_\ell)_i - \frac{C_1 \|\mathrm{Res}_\ell^y(\mu)\|_{V_h'} + \varepsilon_{\tilde{\mathsf{M}}}\,(y_h - y_\ell)}{\tilde{\mathsf{M}}_{ii}^{1/2}}.
\end{aligned}$$

With $\mathsf{g}^{\pm}_{\varepsilon_{\tilde{\mathsf{M}}}}$ as introduced in (15) and the splitting of the nonpositive and positive components of the RB solution, we combine these estimates to finalize the proof because

$$\begin{aligned}
&\big((\Theta(\Psi_\ell \mathsf{y}_\ell) - \Theta(\mathsf{y}_h))\Psi_\ell \mathsf{p}_\ell\big)^T \tilde{\mathsf{M}}\big((\Theta(\Psi_\ell \mathsf{y}_\ell) - \Theta(\mathsf{y}_h))\Psi_\ell \mathsf{p}_\ell\big) \\
&\le \big((\Theta(\mathsf{g}^+_{\varepsilon_{\tilde{\mathsf{M}}}}))\Psi_\ell \mathsf{p}_\ell\big)^T \tilde{\mathsf{M}}(\Theta_1(-\Psi_\ell \mathsf{y}_\ell))\big((\Theta(\mathsf{g}^+_{\varepsilon_{\tilde{\mathsf{M}}}}))\Psi_\ell \mathsf{p}_\ell\big) \\
&\quad + \big((\Theta(\Psi_\ell \mathsf{y}_\ell) - \Theta(\mathsf{g}^-_{\varepsilon_{\tilde{\mathsf{M}}}}))\Psi_\ell \mathsf{p}_\ell\big)^T \tilde{\mathsf{M}}(\Theta(\Psi_\ell \mathsf{y}_\ell))\big((\Theta(\Psi_\ell \mathsf{y}_\ell) - \Theta(\mathsf{g}^-_{\varepsilon_{\tilde{\mathsf{M}}}}))\Psi_\ell \mathsf{p}_\ell\big).
\end{aligned}$$

□

The term $\Delta^p_{w,\varepsilon_{\tilde{\mathsf{M}}}}(V_\ell, \mu)$ still depends on FE solutions, which are costly to compute and, to make matters worse, enter in the complicated grid dependent quadrature error. Its evaluation therefore remains intractable in practice, unless a tight upper

bound on the quadrature error can be derived. We do not provide such a result and do not expect that one is easily obtainable. In the numerical results of the next subsection, we instead assume that the mass lumping quadrature error is small compared to the other error sources in our computation and use the term

$$\begin{aligned}\Delta_w^p(V_\ell,\mu) \coloneqq C_1\Big(&\gamma\Delta_w^y(V_\ell,\mu) + \|\mathrm{Res}_\ell^p(\mu)\|_{V_h'}\\ &+ \Big[\left(\Theta(\mathsf{g}^+)\mathbf{\Psi}_\ell\mathsf{p}_\ell\right)^T \tilde{\mathsf{M}}\Theta_1(-\mathbf{\Psi}_\ell\mathsf{y}_\ell)\left(\Theta(\mathsf{g}^+)\mathbf{\Psi}_\ell\mathsf{p}_\ell\right)\\ &+ \left(\left(\Theta(\mathbf{\Psi}_\ell\mathsf{y}_\ell)-\Theta(\mathsf{g}^-)\right)\mathbf{\Psi}_\ell\mathsf{p}_\ell\right)^T \tilde{\mathsf{M}}\Theta(\mathbf{\Psi}_\ell\mathsf{y}_\ell)\left(\left(\Theta(\mathbf{\Psi}_\ell\mathsf{y}_\ell)-\Theta(\mathsf{g}^-)\right)\mathbf{\Psi}_\ell\mathsf{p}_\ell\right)\Big]^{\frac{1}{2}}\Big),\end{aligned}$$

as an adjoint error indicator, which is the result of assuming that all quadrature errors in $\Delta^p_{w,\varepsilon_{\tilde{\mathsf{M}}}}$ vanish and which is easily evaluable. Interestingly, we obtain the reproduction of the solution results analogously to Corollary 2.6 in the PDE analysis for both the combined error indicators $\Delta^{y,p}_{w,\varepsilon_{\tilde{\mathsf{M}}}}(V_\ell,\mu) = \Delta^y_w(V_\ell,\mu) + \Delta^p_{w,\varepsilon_{\tilde{\mathsf{M}}}}(V_\ell,\mu)$ and $\Delta^{y,p}_w(V_\ell,\mu) = \Delta^y_w(V_\ell,\mu) + \Delta^p_w(V_\ell,\mu)$, which motivates the use of the latter in the numerics.

Corollary 3.5. *Let the assumptions of Lemma* 3.4 *hold and* $\mu \in \mathbb{R}^p$ *with* y_h, $p_h \in V_\ell$. *Then*

1) $y_h = y_\ell$ *and* $p_h = p_\ell$,
2) $\Delta^{y,p}_{w,\varepsilon_{\tilde{\mathsf{M}}}}(V_\ell,\mu) = \Delta^{y,p}_w(V_\ell,\mu) = 0$.

Proof. 1) $y_h = y_\ell$ follows from Corollary 2.6-1). For the adjoint, we use the fact that $y_h = y_\ell$ and $e_p = p_h - p_\ell \in V_\ell$ to conclude that

$$0 \le \langle \nabla e_p, \nabla e_p\rangle_{L^2} + \langle \mathbb{1}_{\{y_\ell>0\}} e_p, e_p\rangle_{L^2} = j'(y_h) - j'(y_\ell) = j'(y_\ell) - j'(y_\ell) = 0,$$

where the first inequality follows from the coercivity of the adjoint equation.

2) The equality $\Delta^y_w(V_\ell,\mu) = 0$ follows from Corollary 2.6-2). We infer from 1) that, since the quadrature error of the constant zero function vanishes, $\Delta^{y,p}_{w,\varepsilon_{\tilde{\mathsf{M}}}}(V_\ell,\mu) = \Delta^{y,p}_w(V_\ell,\mu)$ and additionally $\|\mathrm{Res}^p_\ell(\mu)\|_{V_h'} = 0$. Furthermore, since $\Delta^y_w(V_\ell,\mu) = 0$, we can conclude that $\mathsf{g}^+ = \mathsf{g}^- = \mathbf{\Psi}_\ell\mathsf{y}_\ell$. This implies $\Delta^p_w(V_\ell,\mu) = \Delta^p_{w,\varepsilon_{\tilde{\mathsf{M}}}}(V_\ell,\mu) = 0$ and thus $\Delta^{y,p}_{w,\varepsilon_{\tilde{\mathsf{M}}}}(V_\ell,\mu) = \Delta^{y,p}_w(V_\ell,\mu) = 0$.

□

3.4 Numerical Results for the Optimal Control Problem

Compared to the analytical results for the model order reduction of the PDE constraint in Section 2, there are quite few analytical results on the optimal control problem in the previous subsections. This is largely due to the fact that the indicator function in the adjoint equation essentially prevents the straight-forward derivation of an analytical a-posteriori error estimate. However, with the additional use of the discretization structure, the error indicator derived in Lemma 3.4 offers a way to deal with error indication in the numerics. Furthermore, the parameter space is no longer necessarily compact in the optimization setting, which is typically a fundamental assumption for the offline/online reduced basis method. The adaptive RB approach introduced in Section 3.2 is designed to circumvent this issue. In this subsection, we present numerical results based on the true V-error computation and the error indicator derived in Lemma 3.4, comparing the two approaches for a truly finite dimensional parameter space and showing the promising performance of the adaptive RB method for an inherently infinite dimensional control problem, which usually breaks offline/online RB approaches.

For the remainder of this section, we will assume that $j(y) \coloneqq \frac{1}{2}\|y - y^d\|_{L^2}^2$ and write $\mathsf{B} \in \mathbb{R}^{N\times p}$ for the matrix corresponding to the discretized linear operator $\mathcal{B}\colon \mathbb{R}^p \to V'$. As in Section 2.2, we fix $\Omega = (0,1)^2$ and consider P_1-type finite elements on a Friedrichs-Keller triangulation of the domain in the discretization. We maintain the mass lumping procedure in the nonlinear max term and the indicator function and obtain the following FE and RB stationarity systems from the discretization procedures applied to (13):

$$
\begin{aligned}
\mathsf{K}\bar{\mathsf{y}}_h + \tilde{\mathsf{M}}\max\{0,\bar{\mathsf{y}}_h\} &= \mathsf{B}\bar{\mu} && \text{in } \mathbb{R}^N,\\
\mathsf{K}\bar{\mathsf{p}}_h + \tilde{\mathsf{M}}\Theta(\bar{\mathsf{y}}_h)\bar{\mathsf{p}}_h &= \mathsf{M}(\bar{\mathsf{y}}_h - \mathsf{y}^d) && \text{in } \mathbb{R}^N,\\
\mathsf{B}^T\bar{\mathsf{p}}_h + \sigma\mathsf{A}\bar{\mu} &= 0 && \text{in } \mathbb{R}^p,
\end{aligned}
$$

$$
\begin{aligned}
\Psi_\ell^T\mathsf{K}\Psi_\ell\bar{\mathsf{y}}_\ell + \Psi_\ell^T\tilde{\mathsf{M}}\max\{0,\Psi_\ell\bar{\mathsf{y}}_\ell\} &= \Psi_\ell^T\mathsf{B}\bar{\mu} && \text{in } \mathbb{R}^\ell,\\
\Psi_\ell^T\mathsf{K}\Psi_\ell\bar{\mathsf{p}}_\ell + \Psi_\ell^T\tilde{\mathsf{M}}\Theta(\Psi_\ell\bar{\mathsf{y}}_\ell)\bar{\mathsf{p}}_\ell &= \Psi_\ell^T\mathsf{M}(\Psi_\ell\bar{\mathsf{y}}_\ell - \mathsf{y}^d) && \text{in } \mathbb{R}^\ell,\\
\mathsf{B}^T\Psi_\ell\bar{\mathsf{p}}_\ell + \sigma\mathsf{A}\bar{\mu} &= 0 && \text{in } \mathbb{R}^p.
\end{aligned}
$$

These finite dimensional systems are solved with the PSN method described in the beginning of this section. The FE and RB system matrices at iterates (y_h, p_h, μ),

(y_ℓ, p_ℓ, μ) read as:

$$\begin{pmatrix} \mathsf{K} + \tilde{\mathsf{M}}\Theta(\mathsf{y}_h) & 0 & -\mathsf{B} \\ -\mathsf{M} & \mathsf{K} + \tilde{\mathsf{M}}\Theta(\mathsf{y}_h) & 0 \\ 0 & \mathsf{B}^T & \sigma\mathsf{A} \end{pmatrix},$$
$$\begin{pmatrix} \Psi_\ell^T \mathsf{K}\Psi_\ell + \Psi_\ell^T \tilde{\mathsf{M}}\Theta(\Psi_\ell \mathsf{y}_\ell)\Psi_\ell & 0 & -\Psi_\ell^T \mathsf{B} \\ -\Psi_\ell^T \mathsf{M}\Psi_\ell & \Psi_\ell^T \mathsf{K}\Psi_\ell + \Psi_\ell^T \tilde{\mathsf{M}}\Theta(\Psi_\ell \mathsf{y}_\ell)\Psi_\ell & 0 \\ 0 & \mathsf{B}^T\Psi_\ell & \sigma\mathsf{A} \end{pmatrix}.$$

As for the numerical results in Section 2, our code is implemented in Python3 and uses FENICS [1] for the matrix assembly. Sparse memory management and computations are implemented with SciPy [24]. All computations below were run on an Ubuntu 18.04 notebook with 12 GB main memory and an Intel Core i7-4600U CPU.

Example 3.1. For the first numerical example of this section, we reuse the setting of Example 2.1, i.e., we set $\mathcal{P} = [-2, 2]^2$, $c(\mu) = 1$, $a(\mu) = 10$ (note that this does not change the analysis) and the right hand side as defined in (11). Additionally, we use $y^d(\boldsymbol{x}) = \left(\frac{1}{2} - x_1\right) \sin(\pi x_1) \sin(\pi x_2)$ for $\boldsymbol{x} = (x_1, x_2) \in \Omega$, $\mathsf{A} = \mathsf{I}_2$ the identity in $\mathbb{R}^2$ and $\sigma = 10^{-3}$. As previously, we fix a training set of 121 and a test set of 196 equidistant points in $\mathcal{P}$, where both sets are chosen disjoint. The initial value in all computations is set to $\mu_0 = [1, 1]$. We keep the tolerances fixed at 10^{-4} for the RB error, at 10^{-8} for the semismooth Newton method and now additionally at 10^{-6} for the PSN method. The desired state, stationary state and stationary adjoint computed with finite elements are shown in Figure 4. The results of the

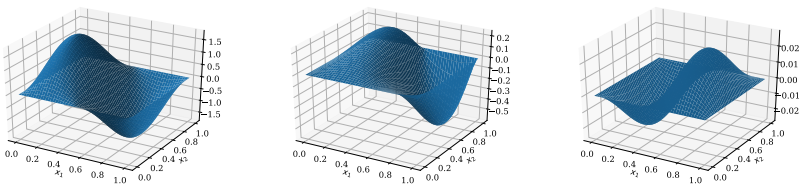

Fig. 4: Example 3.1. Desired state (left), FE stationary state (middle) and FE stationary adjoint (right) for $1/h = 400$.

different approaches for different grids with step size $h = 1/100, 1/200, 1/400$ per dimension can be found in Table 5. For the adaptive RB approach, all parameter combinations $(\eta, n_{\text{fix}}) \in \{0.5, 0.25\} \times \{1, 2, 3, 4\}$ are tested and the best, worst and average performances are stated. First of, we can again note that all of the V-

errors are below the given tolerance of the offline/online and adaptive RB method, respectively. The error indicator's overestimation leads to only slightly smaller errors in the offline/online approach but produce significantly lower errors, larger reduced bases and slower speed-up in the adaptive RB approach, compared to the true error computation. Additionally, the reduced bases for the adaptive RB approach are much smaller than those for the offline/online approach and contain only six or eight elements, depending on the utilized error indicator. This also suggests that a basis of only two more elements is enough to largely reduce the V-errors in the adaptive case. Furthermore, the size of the basis in the adaptive RB seems to be independent of the chosen parameters η and n_{fix} and the step size h in the adaptive RB approach. As before, the number of PSN iterations for solving the FE/RB systems appears to be independent of the step size h and is the same for the full FE system and the offline/online RB approach. Of course, this number differs for the adaptive RB approach with its prescribed number of n_{fix} fixed PSN steps per basis update. In general, the speed-up in the online phase disregarding the offline computational time in the classical RB approach is, as expected, far greater than that of the adaptive RB approach (up to two orders of magnitude). Also as in the previous examples, the true V-error outperforms the error indicator in the RB greedy procedure. It is worth mentioning that for step size $h = 1/400$, the offline/online RB approach with true V-error still requires less total computation time than the FE approach even if the offline time is taken into account. For the adaptive RB approach, the results show significant speed-ups in all cases, which increase with a factor of about three to four when the step size h is divided by two. When we consider best and average performance, computations using the true V-error are always faster compared to the error indicator, but when we consider worst performance the error indicator always slightly outperforms the true V-error. Note that the average speed-up is much closer to the best speed-up than it is to the worst, especially for the true error. Additionally, the worst performance is always achieved for the case $n_{fix} = 1$ and mostly for $\eta = 0.5$, which can be expected, since in this case an evaluation of the (costly) error indicator is necessary in every iteration. Accordingly, choices $n_{fix} \in \{2, 3\}$, where the algorithm reduces the norm of the residual in the new updated RB space considerably before updating the basis further, are arguably more practical.

We conclude that the adaptive RB approach can be suitable to solve the stationarity system and outperform the classical offline/online approach, which is generally a suboptimal choice when only few solves of the problem are required. The new method computes solutions whose approximation quality of the FE solutions can be controlled well at a speed-up of up to 40 using a reduced basis with comparatively few elements.

The performance and effectivity for the error indicator is shown in Table 6. We can observe that the effectivity appears to be mesh dependent here – in contrast to the effectivity of the analytical estimator in Section 2. This is be expected because the indicator includes grid dependent parts and the mass lumping error is ignored. Regardless, even for large h, the performance is reasonable and it improves for finer grids. The lower speed-up of approximately two, compared to the speed-up of four in Section 2, is due to the fact that the adjoint equation is linear. Therefore solving the linear system for the error estimator does not save time compared to solving the adjoint equation. Of course, the computation of the state remains nonlinear – hence the remaining speed-up.

Example 3.2. The second example in this section is designed to show that the adaptive approach presented above can overcome the curse of dimensionality that comes with classical offline/online approaches for large parameter spaces. We consider an FE discretized approximation of the standard optimal control problem with L_2 controls on the right hand side of the PDE – and therefore an inherently infinite dimensional setting. We set $p = N$, $\mathsf{A} = \mathsf{B} = \mathsf{M}$ as the mass matrix and $y^d(x_1, x_2)$ as the solution of (12) in Section 2.2 for $\mu = 0$ – i.e., the function consisting of three bumps on three quarters of the domain. This implies that

$$\bar{p}_h = \frac{1}{\sigma}\bar{\mu}$$

in the FE system. Accordingly, the reduced basis can be used to approximate the parameter space itself because the adjoint and the parameter are linear dependent in the solution. Additionally, we set $\sigma = 10^{-4}$ and for increased influence of the nonsmoooth term, we set $a(\mu) = 100$ as the scaling parameter in front of the max term (note that this does not change the analysis). The different settings for η, n_{fix} and the grids remain the same as they were in Example 3.1. In Figure 5, the desired state, FE stationary state and FE stationary adjoint are shown.

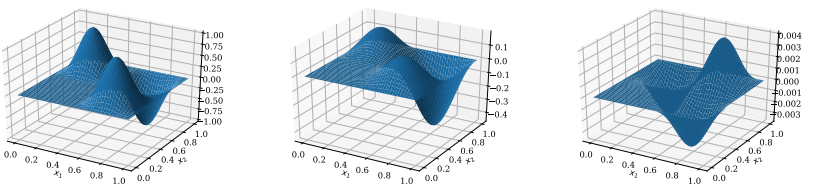

Fig. 5: Example 3.2. Desired state (left), FE stationary state (middle) and FE stationary adjoint (right) for $1/h = 400$.

For the reasons stated above, the offline/online RB approach will not be compared here. Instead, only the adaptive RB approach and the standard FE solution process are compared in Table 7.
First, note that again the number of PSN iterations seems to be mesh-independent in the FE formulations. The iterations in the adaptive approach also show mesh-independence when best performances are compared with best performances of finer/coarser grids, etc. This time, though, the number of (average) iterations for the adaptive RB approach is about twice as large as in the previous example. This is also the case for the size of the reduced basis, which always contains 12 elements for the true error and between 14 and 16 elements for the error indicator. Compared to the previous example, the additional basis elements do not increase the quality of the results significantly. A richer reduced basis was to be expected due to the underlying high dimensional structure of the problem. The resulting speed-ups are not as high as in the previous example and do not increase as quickly as before when h is decreased. While it is reasonable to assume that this is due to the difficulties with the reduced basis approximation, note that the computational time for solving the FE systems does not increase as strongly as before as h is decreased either – compare 765.69 seconds with 4008.98 seconds total in the last example for the finest grid. Nonetheless, a speed-up of close to four can be obtained and the FE setting never outperforms the adaptive RB approach. This shows that the adaptive RB approach can be used as a reasonable approach when RB model order reduction on very high dimensional parameter spaces is desired.

4 Conclusion

We have established an a-posteriori error estimator and presented corresponding numerical considerations for classical offline/online RB approaches for solving a generalized version of the constraining nonsmooth PDE of (**P**) efficiently. The results show promising speed-up while suggesting that the nonsmooth and nonlinear behavior of the max term in the PDE is in fact the part that is most difficult to capture in the RB approximation. For the stationarity system of the optimal control problem (**P**) itself, we have proposed a novel adaptive RB pseudo semismooth Newton approach that creates the required reduced bases adaptively as the PSN algorithm progresses, using local information along the PSN iterates. The adaptive approach is complementary to standard offline/online approaches in the sense that it offers reasonable to large speed-up in online computations without the additional penalty of offline computation times.

FE						
1/h		**it.**	**time (s)**			
100		4	8.96			
200		4	146.28			
400		4	4008.98			
offline/online RB (true V-error)						
1/h		**it.**	**speed-up**	t_{off} **(s)**	$\|e_y\|_V$	$\|e_p\|_V$	$\|\Psi\|$
100		4	70.67	98.03	$2.67 \cdot 10^{-5}$	$3.75 \cdot 10^{-5}$	60
200		4	373.75	523.89	$1.28 \cdot 10^{-5}$	$3.42 \cdot 10^{-5}$	58
400		4	1544.93	3427.74	$7.66 \cdot 10^{-6}$	$3.36 \cdot 10^{-5}$	58
offline/online RB (indicator)						
1/h		**it.**	**speed-up**	t_{off} **(s)**	$\|e_y\|_V$	$\|e_p\|_V$	$\|\Psi\|$
100		4	30.45	1054.56	$1.28 \cdot 10^{-5}$	$2.25 \cdot 10^{-5}$	84
200		4	235.92	6770.66	$4.75 \cdot 10^{-6}$	$2.09 \cdot 10^{-5}$	84
400		4	1105.49	47892.43	$2.09 \cdot 10^{-6}$	$1.98 \cdot 10^{-5}$	84
adaptive RB (true V-error)						
1/h	(η, n_{fix})	**it.**	**speed-up**	$\|e_y\|_V$	$\|e_p\|_V$	$\|\Psi\|$
100	best $(0.25, 3)$	9	3.81	$2.38 \cdot 10^{-5}$	$5.69 \cdot 10^{-6}$	6
	worst $(0.25, 1)$	6	2.14	$2.08 \cdot 10^{-5}$	$5.63 \cdot 10^{-6}$	6
	avg.	8.25	3.35	$2.22 \cdot 10^{-5}$	$5.65 \cdot 10^{-6}$	6
200	best $(0.25, 2)$	6	10.35	$1.96 \cdot 10^{-5}$	$4.44 \cdot 10^{-6}$	6
	worst $(0.5, 1)$	6	5.88	$1.96 \cdot 10^{-5}$	$4.44 \cdot 10^{-6}$	6
	avg.	8.25	9.18	$2.12 \cdot 10^{-5}$	$4.64 \cdot 10^{-6}$	6
400	best $(0.25, 3)$	9	40.87	$2.26 \cdot 10^{-5}$	$4.23 \cdot 10^{-6}$	6
	worst $(0.5, 1)$	6	22.85	$1.94 \cdot 10^{-5}$	$3.83 \cdot 10^{-6}$	6
	avg.	8.25	36.14	$2.09 \cdot 10^{-5}$	$3.90 \cdot 10^{-6}$	6
adaptive RB (indicator)						
1/h	(η, n_{fix})	**it.**	**speed-up**	$\|e_y\|_V$	$\|e_p\|_V$	$\|\Psi\|$
100	best $(0.5, 2)$	12	2.73	$3.87 \cdot 10^{-10}$	$7.35 \cdot 10^{-10}$	8
	worst $(0.5, 1)$	6	2.35	$1.09 \cdot 10^{-8}$	$6.05 \cdot 10^{-8}$	8
	avg.	10.25	2.63	$3.02 \cdot 10^{-9}$	$1.64 \cdot 10^{-8}$	8
200	best $(0.25, 2)$	8	7.22	$7.14 \cdot 10^{-11}$	$4.38 \cdot 10^{-10}$	8
	worst $(0.5, 1)$	6	6.26	$8.73 \cdot 10^{-9}$	$4.86 \cdot 10^{-8}$	8
	avg.	10.25	7.07	$1.35 \cdot 10^{-9}$	$7.81 \cdot 10^{-9}$	8
400	best $(0.25, 2)$	8	28.14	$2.95 \cdot 10^{-11}$	$4.13 \cdot 10^{-10}$	8
	worst $(0.5, 1)$	6	24.38	$2.38 \cdot 10^{-9}$	$2.24 \cdot 10^{-6}$	8
	avg.	10.25	27.50	$4.12 \cdot 10^{-10}$	$4.74 \cdot 10^{-9}$	8

Tab. 5: Example 3.1. The first part of the table shows the number of required pseudo semismooth Newton iterations for computing the FE solution to the stationarity system and the required time. The second and third parts show the performance of the offline/online RB approach including the offline computation time t_{off} The fourth and fifth parts show the best, worst and average performance of the adaptive RB algorithm over all $(\eta, n_{\text{fix}}) \in \{0.5,\ 0.25\} \times \{1, 2, 3, 4\}$ using the true V-error and error estimator, respectively.

1/h	avg. true V-error	avg. effectivity	time true V-error (s)	speed-up indicator
100	$4.85 \cdot 10^{-5}$	36.85	114.36	2.32
200	$4.58 \cdot 10^{-5}$	70.08	697.43	2.27
400	$4.53 \cdot 10^{-5}$	133.18	4933.28	2.18

Tab. 6: Example 3.1. Performance of true error and error indicator with reduced basis from strong greedy on test set.

FE						
1/h		it.	time (s)			
100		5	10.80			
200		5	84.18			
400		5	765.69			
adaptive RB (true V-error)						
1/h	(η, n_{fix})	it.	speed-up	$\|e_y\|_V$	$\|e_p\|_V$	$\|\Psi\|$
100	best $(0.5, 3)$	18	2.63	$2.28 \cdot 10^{-5}$	$9.27 \cdot 10^{-8}$	12
	worst $(0.25, 1)$	9	1.47	$3.27 \cdot 10^{-5}$	$4.79 \cdot 10^{-7}$	12
	avg.	16	2.14	$2.22 \cdot 10^{-5}$	$1.37 \cdot 10^{-7}$	12
200	best $(0.25, 3)$	18	2.90	$1.41 \cdot 10^{-5}$	$8.47 \cdot 10^{-7}$	12
	worst $(0.5, 1)$	9	1.73	$1.74 \cdot 10^{-5}$	$8.70 \cdot 10^{-7}$	12
	avg.	15.87	2.43	$1.35 \cdot 10^{-5}$	$5.01 \cdot 10^{-7}$	12
400	best $(0.25, 4)$	24	3.92	$1.02 \cdot 10^{-5}$	$5.24 \cdot 10^{-8}$	12
	worst $(0.5, 1)$	9	2.32	$9.79 \cdot 10^{-6}$	$6.47 \cdot 10^{-8}$	12
	avg.	15.87	3.31	$1.12 \cdot 10^{-5}$	$1.13 \cdot 10^{-7}$	12
adaptive RB (indicator)						
1/h	(η, n_{fix})	it.	speed-up	$\|e_y\|_V$	$\|e_p\|_V$	$\|\Psi\|$
100	best $(0.5, 3)$	21	1.88	$4.11 \cdot 10^{-7}$	$3.05 \cdot 10^{-9}$	14
	worst $(0.5, 1)$	10	1.51	$1.58 \cdot 10^{-6}$	$7.14 \cdot 10^{-9}$	16
	avg.	18.12	1.72	$1.17 \cdot 10^{-6}$	$7.84 \cdot 10^{-9}$	14.25
200	best $(0.2, 3)$	21	2.06	$1.01 \cdot 10^{-5}$	$8.55 \cdot 10^{-7}$	14
	worst $(0.5, 1)$	9	1.88	$1.09 \cdot 10^{-5}$	$8.55 \cdot 10^{-7}$	14
	avg.	18	1.99	$9.62 \cdot 10^{-6}$	$8.56 \cdot 10^{-7}$	14
400	best $(0.25, 3)$	21	2.76	$2.57 \cdot 10^{-7}$	$1.17 \cdot 10^{-9}$	14
	worst $(0.5, 1)$	10	2.27	$1.41 \cdot 10^{-6}$	$6.47 \cdot 10^{-9}$	16
	avg.	18.12	2.64	$4.98 \cdot 10^{-6}$	$4.56 \cdot 10^{-8}$	14.25

Tab. 7: Example 3.2. The first part of the table shows the number of required PSN iterations for computing the FE solution to the stationarity system and the required time. The second and third parts show the best, worst and average performance of the adaptive RB algorithm over all $(\eta, n_{\mathrm{fix}}) \in \{0.5,\ 0.25\} \times \{1, 2, 3, 4\}$ using the true V-error and error indicator, respectively.

Acknowledgment: This research was supported by the German Research Foundation (DFG) under grant number VO 1658/5-2 within the priority program "Non-smooth and Complementarity-based Distributed Parameter Systems: Simulation and Hierarchical Optimization" (SPP 1962).

References

[1] M. Alnæs, J. Blechta, J. Hake, A. Johansson, B. Kehlet, A. Logg, C. Richardson, J. Ring, M.E. Rognes, and G.N. Wells. The fenics project version 1.5. *Archive of Numerical Software*, 3(100), 2015.

[2] M. Barrault, Y. Maday, N.C. Nguyen, and A.T Patera. An empirical interpolation method: application to efficient reduced-basis discretization of partial differential equations. *Comptes Rendus Mathematique*, 339(9):667–672, 2004.

[3] P. Benner, M. Ohlberger, A. Cohen, and K. Willcox. *Model Reduction and Approximation: Theory and Algorithms*. Computational Science and Engineering. SIAM, 2017.

[4] M. Bernreuther. Rb-based pde-constrained non-smooth optimization. Master's thesis, Universität Konstanz, 2019.

[5] C. Canuto, T. Tonn, and K. Urban. A-posteriori error analysis of the reduced basis method for non-affine parameterized nonlinear pde's. *SIAM Journal on Numerical Analysis*, 47(3):2001–2022, 2009.

[6] S. Chaturantabut and D.C. Sorensen. Nonlinear model reduction via discrete empirical interpolation. *SIAM Journal on Scientific Computing*, 32(5):2737–2764, 2010.

[7] S. Chaturantabut and D.C. Sorensen. A state space estimate for POD-DEIM nonlinear model reduction. *SIAM Journal on Numerical Analysis*, 50(1):46–63, 2012.

[8] C. Christof, C. Clason, C. Meyer, and S. Walther. Optimal control of a non-smooth semilinear elliptic equation. *Mathematical Control and Related Fields (MCRF)*, 8:247–276, 2018.

[9] C. Christof and G. Müller. Multiobjective optimal control of a non-smooth semilinear elliptic partial differential equation. Technical report, 2020. Submitted.

[10] D. Gilbarg and N.S. Trudinger. *Elliptic Partial Differential Equations of Second Order*. Classics in Mathematics. Springer, 2001.

[11] M.A. Grepl, Y. Maday, N.C. Nguyen, and A.T. Patera. Efficient reduced-basis treatment of nonaffine and nonlinear partial differential equations. *ESAIM: Mathematical Modelling and Numerical Analysis*, 41(3):575–605, 2007.

[12] B. Haasdonk. A tutorial on rb-methods. In P. Benner, A. Cohen, M. Ohlberger, and K. Willcox, editors, *Model Reduction and Approximation: Theory and Algorithms*, Computational Science & Engineering, pages 67–138. SIAM, 2017.

[13] J. Hesthaven, G. Rozza, and B. Stamm. *Certified Reduced Basis Methods for Parametrized Partial Differential Equations*. SpringerBriefs in Mathematics. Springer, 2016.

[14] M. Hinze and D. Korolev. Reduced basis methods for quasilinear elliptic pdes with applications to permanent magnet synchronous motors. Technical report, 2020. Submitted.

[15] M. Kärcher, Z. Tokoutsi, M. Grepl, and K. Veroy. Certified reduced basis methods for parametrized elliptic optimal control problems with distributed controls. *Journal of Scientific Computing*, 75:276–307, 2018.

[16] F. Kikuchi, K. Nakazato, and T. Ushijima. Finite element approximation of a nonlinear eigenvalue problem related to MHD equilibria. *Japan J. Appl. Math.*, 1(2):369–403, 1984.

[17] F. Negri, G. Rozza, A. Manzoni, and A. Quarteroni. Reduced basis method for parametrized elliptic optimal control problems. *SIAM Journal on Scientific Computing*, 35:A2316–A2340, 2013.

[18] A.T. Patera and G. Rozza. *Reduced Basis Approximation and A Posteriori Error Estimation for Parametrized Partial Differential Equations*. MIT Pappalardo Graduate Monographs in Mechanical Engineering, 2007.

[19] E. Qian, M. Grepl, K. Veroy, and K. Willcox. A certified trust region reduced basis approach to PDE-constrained optimization. *SIAM Journal on Scientific Computing*, 39(5):S434–S460, 2017.

[20] A. Quarteroni, A. Manzoni, and F. Negri. *Reduced Basis Methods for Partial Differential Equations*. Unitext Series. Springer, 2016.

[21] J. Rappaz. Approximation of a nondifferentiable nonlinear problem related to MHD equilibria. *Numer. Math.*, 45(1):117–133, 1984.

[22] R. Temam. A non-linear eigenvalue problem: the shape at equilibrium of a confined plasma. *Arch. Rational Mech. Anal.*, 60(1):51–73, 1976.

[23] K. Veroy, C. Prud'homme, D.V. Rovas, and A.T. Patera. A posteriori error bounds for reduced-basis approximation of parametrized noncoercive and nonlinear elliptic partial differential equations. In *16th AIAA Computational Fluid Dynamics Conference, 2003, Orlando, United States*, 2003.

[24] P. Virtanen, R. Gommers, T.E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S.J. van der Walt, M. Brett, J. Wilson, K. Jarrod Millman, N. Mayorov, A.R.J. Nelson, E. Jones, R. Kern, E. Larson, C.J. Carey, İ. Polat, Y. Feng, E.W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E.A. Quintero, C.R. Harris, A.M. Archibald, A.H. Ribeiro, F. Pedregosa, P. van Mulbregt, and SciPy 1. 0

Contributors. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272, 2020.

[25] J. Xin. *An Introduction to Fronts in Random Media*, volume 5 of *Surveys and Tutorials in the Applied Mathematical Sciences*. Springer, 2009.