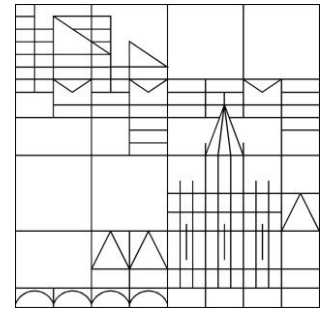


Universität Konstanz



Optimality conditions and POD a-posteriori error estimates for a
semilinear parabolic optimal control

Oliver Lass
Stefan Trenz
Stefan Volkwein

Konstanzer Schriften in Mathematik

Nr. 345, November 2015

ISSN 1430-3558

Konstanzer Online-Publikations-System (KOPS)
URL: <http://nbn-resolving.de/urn:nbn:de:bsz:352-0-310049>

OPTIMALITY CONDITIONS AND POD A-POSTERIORI ERROR ESTIMATES FOR A SEMILINEAR PARABOLIC OPTIMAL CONTROL

O. LASS, S. TRENZ, AND S. VOLKWEIN

ABSTRACT. In the present paper the authors consider an optimal control problem for a parametrized nonlinear parabolic differential equation, which is motivated by lithium-ion battery models. A standard finite element (FE) discretization leads to a large-scale nonlinear optimization problem so that its numerical solution is very costly. Therefore, a reduced-order modelling based on proper orthogonal decomposition (POD) is applied, so that the number of degrees of freedom is reduced significantly and a fast numerical simulation of the model is possible. To control the error, an a-posteriori error estimator is realized. Numerical experiments show the efficiency of the approach.

1. INTRODUCTION

We consider an optimal control problem which is governed by a semilinear parabolic partial differential equation (PDE) and bilateral control constraints. The PDE occurs in lithium-ion battery models (see [8, 28]) as an equation for the concentration of lithium-ions. This equation describes the mass transport in the (positive) electrode of a battery. Notice that the modeling and optimization of lithium-ion batteries has received an increasing amount of attention in the recent past.

The discretization of the nonlinear optimal control problem using, e.g., finite element techniques, lead to very large discretized systems that are expensive to solve. The goal is to develop a reduced order model for the nonlinear system of PDEs that is cheap to evaluate. In this paper we apply the method of proper orthogonal decomposition (POD); see, e.g., [13, 17, 25]. POD is used to generate a basis of a subspace that expresses the characteristics of the expected solution. This is in contrast to more general approximation methods, such as the finite element method, that do not correlate to the dynamics of the underlying system. We refer the reader to [3], where the authors apply a POD reduced-order modeling for the lithium ion battery model presented in [8].

When using a reduced order model in the optimization process an error is introduced. Therefore, an a posteriori error estimator has to be developed in order to quantify the quality of the obtained solution. We here use recent results from

Date: November 26, 2015.

2000 Mathematics Subject Classification. 49K20, 65K10, 65K30.

Key words and phrases. Optimal control, semilinear parabolic differential equations, a-posteriori analysis, proper orthogonal decomposition, projected Newton method.

The first author gratefully acknowledges support by the German Science Fund *Numerical and Analytical Methods for Elliptic-Parabolic Systems Appearing in the Modeling of Lithium-Ion Batteries* (Excellence Initiative). The second author gratefully acknowledges support by the German Science Fund *A-Posteriori-POD Error Estimates for Nonlinear Optimal Control Problems governed by Partial Differential Equations*.

[6, 15, 20]. Further, it is important to understand that the obtained reduced order model by the POD is only a local approximation of the nonlinear system. Hence, it is necessary to guarantee that the approximation is good throughout the optimization process. For this we make use of a simplified error indicator known from the reduced basis strategies [9]. Using this error indicator an adaptive reduced order model strategy is proposed to solve the optimization problem in an efficient way.

However, to obtain the state data underlying the POD reduced order model, it is necessary to solve once the full state system and consequently the POD approximations depend on the chosen parameters for this solve. To be more precise, the choice of an initial control turned out to be essential. When using an arbitrary control, the obtained accuracy was not at all satisfying even when using a huge number of basis functions whereas an optimal POD basis (computed from the FE optimally controlled state) led to far better results [12]. To overcome this problem different techniques for improving the POD basis have been proposed; see [2, 18]. We follow a residual based approach, which leads to an adaptive algorithm, which does not require any offline computations.

The paper is organized in the following manner: In Section 2 we formulate the nonlinear, nonconvex optimal control problem. First- and second-order optimality conditions are studied in Section 3. The a-posteriori error analysis is reviewed in Section 4. Then, Section 5 is devoted to the POD method and the reduced-order modeling for the optimal control problem. In Section 6 we present the numerical solution method and the POD basis update regarding the a-posteriori error and the reduced-order residuals. Finally, numerical experiments are presented in Section 7.

2. THE OPTIMAL CONTROL PROBLEM

In this section we formulate the optimal control problem, study the semilinear parabolic state equation, prove the existence of optimal solutions and introduce a so-called reduced optimal control problem.

2.1. The problem formulation. Suppose that $\Omega \subset \mathbb{R}^d$, $d \in \{1, 2, 3\}$ is a bounded and open set with a Lipschitz-continuous boundary $\Gamma = \partial\Omega$. For $T > 0$ we set $Q = (0, T) \times \Omega$ and $\Sigma = (0, T) \times \Gamma$. We set $V = H^1(\Omega)$ and $H = L^2(\Omega)$. For the definition of Sobolev spaces we refer, e.g., to [1, 7]. The space $L^2(0, T; V)$ stands for the space of (equivalence classes) of measurable abstract functions $\varphi : [0, T] \rightarrow V$, which are square integrable, i.e.,

$$\int_0^T \|\varphi(t)\|_V^2 dt < \infty.$$

When t is fixed, the expression $\varphi(t)$ stands for the function $\varphi(t, \cdot)$ considered as a function in Ω only. Recall that

$$W(0, T) = \{\varphi \in L^2(0, T; V) \mid \varphi_t \in L^2(0, T; V)\}$$

is a Hilbert space supplied with its common inner product; see [5, p. 472-479]. We define the bounded, closed and convex parameter set

$$\mathcal{D}_{ad} = \{\mu \in \mathbb{R}^m \mid \mu^a \leq \mu \leq \mu^b\},$$

where $\mu^a = (\mu_1^a, \dots, \mu_m^a)$, $\mu^b = (\mu_1^b, \dots, \mu_m^b) \in \mathbb{R}^m$ satisfy $\mu_i^a \leq \mu_i^b$ and ' \leq ' is interpreted componentwise.

Let us consider the following optimal control problem

$$(2.1a) \quad \min J(y, \mu) = \frac{1}{2} \int_0^T \int_{\Omega} \alpha_Q |y - y_Q|^2 \, d\mathbf{x} dt + \frac{1}{2} \int_{\Omega} \alpha_{\Omega} |y(T) - y_{\Omega}|^2 \, d\mathbf{x} \\ + \frac{\lambda}{2} \sum_{i=1}^m \kappa_i |\mu_i - \mu_i^{\circ}|^2$$

subject to the semilinear parabolic differential equation

$$(2.1b) \quad y_t - \Delta y + \sinh \left(y \sum_{i=1}^m \mu_i b_i \right) = f \quad \text{in } Q,$$

$$(2.1c) \quad \frac{\partial y}{\partial n} = 0 \quad \text{on } \Sigma,$$

$$(2.1d) \quad y(0) = y_{\circ} \quad \text{in } \Omega$$

and the inequality constraint

$$(2.1e) \quad \mu \in \mathcal{D}_{ad}.$$

Throughout we will utilize the following hypotheses.

- Assumption 1.** a) *The set of admissible parameters is given by $\mathcal{D}_{ad} = \{\mu \in \mathbb{R}^m \mid \mu^a \leq \mu \leq \mu^b\}$ with $0 \leq \mu^a \leq \mu^b$; The data of the state equation satisfies $b_1, \dots, b_m \in L^{\infty}(\Omega)$ with $b_i \geq 0$ almost everywhere (a.e.) in Ω , $f \in L^r(Q)$ with $r > d/2 + 1$ and $y_{\circ} \in L^{\infty}(\Omega)$.*
- b) *For the cost function we assume that the weighting functions $\alpha_Q \in L^{\infty}(Q)$, $\alpha_{\Omega} \in L^{\infty}(\Omega)$ are non-negative, that the desired states satisfy $y_Q \in L^{\infty}(Q)$, $y_{\Omega} \in L^{\infty}(\Omega)$, the regularization parameters $\kappa_1, \dots, \kappa_m, \lambda$ are nonnegative scalars and the nominal parameter $\mu^{\circ} = (\mu_1^{\circ}, \dots, \mu_m^{\circ})$ belongs to \mathbb{R}^m .*

2.2. The semilinear parabolic equation. Let us introduce the nonlinear function $d : \Omega \times \mathbb{R}^m \times \mathcal{D}_{ad} \rightarrow \mathbb{R}$ by

$$(2.2) \quad d(\mathbf{x}, y, \mu) = \sinh \left(y \sum_{i=1}^m \mu_i b_i(\mathbf{x}) \right), \quad (\mathbf{x}, y, \mu) \in \Omega \times \mathbb{R} \times \mathcal{D}_{ad}.$$

For fixed $\mu \in \mathbb{R}^m$ we consider the Nemytskii operator $\Phi_{\mu} : L^{\infty}(Q) \rightarrow L^{\infty}(Q)$ given by

$$(\Phi_{\mu}(y))(t, \mathbf{x}) = d(\mathbf{x}, y(t, \mathbf{x}), \mu) \quad \text{for almost all (f.a.a.) } (t, \mathbf{x}) \in [0, T] \times \Omega.$$

It follows by the same arguments as in [26, Lemma 4.12] that the mapping Φ_{μ} is twice continuously Fréchet-differentiable from $L^{\infty}(Q)$ to $L^{\infty}(Q)$ and we have

$$\begin{aligned} (\Phi'_{\mu}(y)y_{\delta})(t, \mathbf{x}) &= d_y(\mathbf{x}, y(t, \mathbf{x}), \mu)y_{\delta}(t, \mathbf{x}) \\ &= \sum_{j=1}^m \mu_j b_j(\mathbf{x}) \cosh \left(y(t, \mathbf{x}) \sum_{i=1}^m \mu_i b_i(\mathbf{x}) \right) y_{\delta}(t, \mathbf{x}), \\ (\Phi''_{\mu}(y)(y_{\delta}, \tilde{y}_{\delta}))(t, \mathbf{x}) &= d_{yy}(\mathbf{x}, y(t, \mathbf{x}), \mu)y_{\delta}(t, \mathbf{x})\tilde{y}_{\delta}(t, \mathbf{x}) \\ &= \left(\sum_{j=1}^m \mu_j b_j(\mathbf{x}) \right)^2 \sinh \left(y(t, \mathbf{x}) \sum_{i=1}^m \mu_i b_i(\mathbf{x}) \right) y_{\delta}(t, \mathbf{x})\tilde{y}_{\delta}(t, \mathbf{x}) \end{aligned}$$

for all $y, y_{\delta}, \tilde{y}_{\delta} \in L^{\infty}(Q)$ and f.a.a. $(t, \mathbf{x}) \in [0, T] \times \Omega$.

Definition 2.1. A function $y \in W(0, T) \cap L^\infty(Q)$ is called a weak solution to (2.1b)–(2.1d) provided $y(0) = y_\circ$ holds and

$$(2.3) \quad \langle y_t(t), \varphi \rangle_{V', V} + \int_{\Omega} \nabla y(t) \cdot \nabla \varphi + d(\cdot, y(t), \mu) \varphi \, d\mathbf{x} = \int_{\Omega} f(t) \varphi \, d\mathbf{x}$$

is satisfied for all $\varphi \in V$ and f.a.a. $t \in [0, T]$, where d is given by (2.2).

We define the state space

$$\mathcal{Y} = W(0, T) \cap L^\infty(Q)$$

endowed with the norm $\|y\|_{\mathcal{Y}} = \|y\|_{W(0, T)} + \|y\|_{L^\infty(Q)}$. The next result follows from Theorem 5.5 in [26, p. 213] and the subsequent remark. We also refer to [4] and [24] for a detailed proof.

Theorem 2.2. Let Assumption 1-a) hold. Then, for any $\mu \in \mathcal{D}_{ad}$ there exists a unique weak solution $y = y(\mu)$ to (2.1b)–(2.1d). Moreover, there exists a constant $C > 0$ independent of μ, f, y_\circ (but dependent on μ^a and μ^b) such that

$$(2.4) \quad \|y\|_{\mathcal{Y}} \leq C (\|f\|_{L^r(Q)} + \|y_\circ\|_{L^\infty(\Omega)}).$$

Remark 2.3. 1) The proof of Theorem 2.2 relies essentially on properties of our nonlinearity (2.2). Since the b_i 's are essentially bounded in Ω , the mapping $d(\cdot, y, \mu) : \Omega \rightarrow \mathbb{R}$ is measurable and essentially bounded in Ω for any fixed $(y, \mu) \in \mathbb{R} \times \mathcal{D}_{ad}$. Moreover, $d(\mathbf{x}, 0, \mu) = 0$ holds, the mapping $y \mapsto d(\mathbf{x}, y, \mu)$ is strictly monotonically increasing and (at least) twice continuously differentiable (i.e., d_y and d_{yy} are locally Lipschitz-continuous) f.a.a. $\mathbf{x} \in \Omega$ and for all $\mu \in \mathbb{R}^m$.

2) From $\mu^a \geq 0$ and $b_i \geq 0$ in Ω a.e. for $1 \leq i \leq m$ we infer that

$$\begin{aligned} d_y(\mathbf{x}, y, \mu) &= \sum_{j=1}^m \mu_j b_j(\mathbf{x}) \cosh \left(y \sum_{i=1}^m \mu_i b_i(\mathbf{x}) \right) \geq 0, \\ d_{yy}(\mathbf{x}, y, \mu) &= \left(\sum_{j=1}^m \mu_j b_j(\mathbf{x}) \right)^2 \sinh \left(y \sum_{i=1}^m \mu_i b_i(\mathbf{x}) \right) \end{aligned}$$

f.a.a. $\mathbf{x} \in \Omega$ and for all $(y, \mu) \in \mathbb{R} \times \mathcal{D}_{ad}$. Further, we have

$$|d_y(\mathbf{x}, 0, \mu)| = \left| \sum_{i=1}^m \mu_i b_i(\mathbf{x}) \right| \leq m \max_{1 \leq i \leq m} \{ \mu_i^b \|b_i\|_{L^\infty(\Omega)} \} =: K,$$

and $|d_{yy}(\mathbf{x}, 0, \mu)| = 0$ f.a.a. $\mathbf{x} \in \Omega$ and for all $\mu \in \mathcal{D}_{ad}$.

3) Since $y_\circ \in L^\infty(\Omega)$ holds, we only get that the weak solution y to (2.1b)–(2.1d) belongs to $C((0, T] \times \Omega)$. If $y_\circ \in C(\bar{\Omega})$ is fulfilled, we even have $y \in C(\bar{Q})$; see [4, 24]. \diamond

Motivated by Theorem 2.2 we introduce the solution operator $\mathfrak{G} : \mathcal{D}_{ad} \rightarrow \mathcal{Y}$, where $y = \mathfrak{G}(\mu)$ is the weak solution to (2.1b)–(2.1d) for given $\mu \in \mathcal{D}_{ad}$. The next result is proved in the Appendix.

Proposition 2.4. Assume that Assumption 1-a) holds. Then, the solution operator \mathfrak{G} is globally Lipschitz-continuous.

2.3. Existence of optimal controls. Our optimal control problem is

$$(P) \quad \min J(y, \mu) \quad \text{subject to} \quad (y, \mu) \in \mathcal{F}(\mathbf{P}),$$

where we define the feasible set by

$$\mathcal{F}(\mathbf{P}) = \{(y, \mu) \in \mathcal{X}_{ad} \mid y \text{ is a weak solution to (2.1b)-(2.1d) for } \mu\}$$

with $\mathcal{X}_{ad} = \mathcal{Y} \times \mathcal{D}_{ad}$. Next we turn to the existence of optimal solutions to (P).

Theorem 2.5. *Let Assumption 1 hold. Then, (P) has at least one (global) optimal solution $\bar{x} = (\bar{y}, \bar{\mu})$.*

Proof. Note that the cost functional is nonnegative. Moreover, $\mathcal{D}_{ad} \neq \emptyset$ holds. Let $\tilde{\mu} \in \mathcal{D}_{ad}$ be chosen arbitrarily and $\tilde{y} = y(\tilde{\mu})$ the corresponding weak solution to (2.1b)-(2.1d). Then we have

$$0 \leq \inf \{J(y, \mu) \mid (y, \mu) \in \mathcal{F}(\mathbf{P})\} \leq J(\tilde{y}, \tilde{\mu}) < \infty.$$

Let $\{(y^n, \mu^n)\}_{n \in \mathbb{N}}$ denote a minimizing sequence in $\mathcal{F}(\mathbf{P})$ for the cost functional J , i.e., $\lim_{n \rightarrow \infty} J(y^n, \mu^n) = \inf \{J(y, \mu) \mid (y, \mu) \in \mathcal{F}(\mathbf{P})\}$. Since \mathcal{D}_{ad} is bounded, there exists a bounded sequence $\{\mu^n\}_{n \in \mathbb{N}}$ in \mathbb{R}^m . Moreover, \mathcal{D}_{ad} is closed. Hence, there is a subsequence, again denoted by $\{\mu^n\}_{n \in \mathbb{N}}$, and an element $\bar{\mu} \in \mathcal{D}_{ad}$ so that

$$(2.5) \quad \mu^n \rightarrow \bar{\mu} \text{ in } \mathbb{R}^m \text{ as } n \rightarrow \infty.$$

Due to (2.5) we have

$$\left\| \sum_{i=1}^m \mu_i^n b_i - \sum_{i=1}^m \bar{\mu}_i b_i \right\|_{L^\infty(\Omega)} \rightarrow 0 \text{ as } n \rightarrow \infty.$$

Thus, there exists a constant $C_1 > 0$ satisfying

$$(2.6) \quad \left\| \sum_{i=1}^m \mu_i^n b_i \right\|_{L^\infty(\Omega)} \leq C_1.$$

From (2.4) and (2.5) we infer the existence of a constant $C_2 > 0$ independent of n satisfying

$$(2.7) \quad \|y^n\|_{L^\infty(Q)} \leq C_2 (\|f\|_{L^r(Q)} + \|y_0\|_{L^\infty(\Omega)}).$$

Using (2.6) and (2.7) we derive that there is a constant $C_3 > 0$ such that

$$(2.8) \quad \left\| \cosh \left(s y^n \sum_{i=1}^m \mu_i^n b_i \right) \right\|_{L^\infty(Q)} \leq C_3 \quad \text{for every } s \in [0, 1].$$

We define the sequence $\{z^n\}_{n \in \mathbb{N}}$ by

$$z^n(t, \mathbf{x}) = -\sinh \left(y^n(t, \mathbf{x}) \sum_{i=1}^m \mu_i^n b_i(\mathbf{x}) \right) \quad \text{f.a.a. } (t, \mathbf{x}) \in Q.$$

Applying the mean value theorem we have

$$\begin{aligned} |z^n(t, \mathbf{x})| &= \left| \sinh \left(y^n(t, \mathbf{x}) \sum_{i=1}^m \mu_i^n b_i(\mathbf{x}) \right) \right| \\ &= \left| \sinh \left(y^n(t, \mathbf{x}) \sum_{i=1}^m \mu_i^n b_i(\mathbf{x}) \right) - \sinh \left(0 \cdot \sum_{i=1}^m \mu_i^n b_i(\mathbf{x}) \right) \right| \\ &= \left| \cosh \left(s_n y^n(t, \mathbf{x}) \sum_{i=1}^m \mu_i^n b_i(\mathbf{x}) \right) \right| |y^n(t, \mathbf{x}) - 0| \end{aligned}$$

f.a.a. $(t, \mathbf{x}) \in Q$ with $s_n \in (0, 1)$. Thus, from (2.8), (2.4) and (2.7) it follows that

$$|z^n(t, \mathbf{x})| \leq C_2 C_3 \left(\|f\|_{L^r(Q)} + \|y_\circ\|_{L^\infty(\Omega)} \right)$$

f.a.a. $(t, \mathbf{x}) \in Q$. Consequently, we can assume that $\{z^n\}_{n \in \mathbb{N}}$ contains a subsequence, again denoted by $\{z^n\}_{n \in \mathbb{N}}$, which converges weakly to an element $z \in L^r(Q)$. Now we can use the same arguments as in the proof of Theorem 5.7 in [26] to derive the existence of an element $\bar{y} \in L^\infty(Q)$ so that

$$(2.9) \quad y^n \rightarrow \bar{y} \text{ in } L^\infty(Q) \text{ as } n \rightarrow \infty.$$

From (2.5) and (2.9) we infer that

$$\sinh \left(y^n \sum_{i=1}^m \mu_i^n b_i \right) \rightarrow \sinh \left(\bar{y} \sum_{i=1}^m \bar{\mu}_i b_i \right) \text{ in } L^\infty(Q) \text{ as } n \rightarrow \infty.$$

Next we consider (2.1b)-(2.1d) in the form

$$\begin{aligned} y_t^n - \Delta y^n &= z^n + f && \text{in } Q, \\ \frac{\partial y^n}{\partial n} &= 0 && \text{on } \Sigma, \\ y^n(0) &= y_\circ && \text{in } \Omega. \end{aligned}$$

Now we can proceed as in the proof of Theorem 5.7 in [26] to prove the claim. Recall that in our case the control variable μ belongs to a finite-dimensional set. \square

2.4. The reduced control problem. Using the parameter-to-state operator $\mathcal{G} : \mathcal{D}_{ad} \rightarrow \mathcal{Y}$ we define the reduced cost functional as

$$\hat{J}(\mu) = J(\mathcal{G}(\mu), \mu) \quad \text{for } \mu \in \mathcal{D}_{ad}.$$

Then, **(P)** is equivalent to the reduced problem

$$(\hat{\mathbf{P}}) \quad \min \hat{J}(\mu) \quad \text{s.t. } \mu \in \mathcal{D}_{ad}.$$

It follows from Theorem 2.5 that $(\hat{\mathbf{P}})$ possesses at least a (global) solution $\bar{\mu} \in \mathcal{D}_{ad}$. Moreover, the pair $\bar{x} = (\mathcal{G}(\bar{\mu}), \bar{\mu})$ is a local solution to **(P)**.

3. FIRST- AND SECOND-ORDER OPTIMALITY CONDITIONS

To characterize a local optimal solution to **(P)** or $(\hat{\mathbf{P}})$ we derive optimality conditions. Our approach is based on the formal Lagrange technique as it is described in [26]. Let us introduce the Lagrange function \mathcal{L} associated with **(P)** considering the weak formulation (2.3)

$$\begin{aligned} \mathcal{L}(y, \mu, p) &= \frac{1}{2} \int_0^T \int_\Omega \alpha_Q |y - y_Q|^2 \, d\mathbf{x} dt + \frac{1}{2} \int_\Omega \alpha_\Omega |y(T) - y_\Omega|^2 \, d\mathbf{x} \\ &\quad + \frac{\lambda}{2} \sum_{i=1}^m \kappa_i |\mu_i - \mu_i^\circ|^2 + \int_0^T \langle y_t(t), p(t) \rangle_{V', V} \, dt \\ &\quad + \int_0^T \int_\Omega \nabla y(t) \cdot \nabla p(t) + (d(\cdot, y(t), \mu) - f(t)) p(t) \, d\mathbf{x} dt \end{aligned}$$

for $(y, \mu, p) \in \mathcal{Y} \times \mathbb{R}^m \times L^2(0, T; V)$, where the nonlinearity d has been introduced in (2.2). Notice that the initial condition and the inequality constraints for μ are not eliminated by introducing corresponding Lagrange multipliers.

3.1. First-order necessary optimality conditions. Let Assumption 1 be satisfied. Suppose that $\bar{\mu} \in \mathcal{D}_{ad}$ is a local optimal solution to $(\hat{\mathbf{P}})$ and $\bar{y} = \mathcal{G}(\bar{\mu})$ the associated optimal state. To derive first-order optimality conditions we have to differentiate the reduced cost functional \hat{J} with respect to the parameter μ . Hence, we must compute the derivative of the mapping \mathcal{G} . Proposition 2.4 implies the following proposition. For a proof we refer the reader to the Appendix.

Proposition 3.1. *Let Assumption 1-a) hold. Then, the operator \mathcal{G} is Fréchet differentiable. Let $\tilde{y} = \mathcal{G}(\tilde{\mu})$ and $y = \mathcal{G}'(\tilde{\mu})\mu$ for $\tilde{\mu} \in \mathcal{D}_{ad}$ and $\mu \in \mathbb{R}^m$. Then, $y \in \mathcal{Y}$ is the weak solution to the linear parabolic problem*

$$(3.1) \quad \begin{aligned} y_t - \Delta y + d_y(\cdot, \tilde{y}(\cdot), \tilde{\mu})y &= -d_\mu(\cdot, \tilde{y}(\cdot), \tilde{\mu})\mu && \text{in } Q, \\ \frac{\partial y}{\partial n} &= 0 && \text{on } \Sigma, \\ y(0) &= 0 && \text{in } \Omega, \end{aligned}$$

where d_y is given in Remark 2.3-2) and the row vector $d_\mu(\cdot, \tilde{y}(\cdot), \tilde{\mu})$ has the components

$$d_{\mu_i}(\mathbf{x}, \tilde{y}(t, \mathbf{x}), \tilde{\mu}) = \tilde{y}(t, \mathbf{x}) \cosh\left(\tilde{y}(t, \mathbf{x}) \sum_{j=1}^m \tilde{\mu}_j b_j(\mathbf{x})\right) b_i(\mathbf{x})$$

for $1 \leq i \leq m$ and f.a.a. $(t, \mathbf{x}) \in Q$. Furthermore,

$$(3.2) \quad \|y\|_{\mathcal{Y}} \leq \bar{C} |\mu|_2$$

with a constant $\bar{C} > 0$ depending in T , $|\Omega|$, \tilde{y} , $\tilde{\mu}$, m and the b_i 's. In (3.2) we denote by $|\cdot|_2$ the Euclidean norm.

Remark 3.2. It follows from (3.2) that the linear operator $\mathcal{G}'(\bar{\mu})$ is bounded. \diamond

Next we consider the following first-order conditions (see, e.g., [11, 26]). For that purpose we study

$$(3.3a) \quad \mathcal{L}_y(\bar{y}, \bar{\mu}, \bar{p})y = 0 \quad \text{for all } y \in \mathcal{Y} \text{ with } y(0) = 0,$$

$$(3.3b) \quad \mathcal{L}_\mu(\bar{y}, \bar{\mu}, \bar{p})(\mu - \bar{\mu}) \geq 0 \quad \text{for all } \mu \in \mathcal{D}_{ad}.$$

The Fréchet derivative $\mathcal{L}_y(\bar{y}, \bar{\mu}, \bar{p})$ in a direction $y \in \mathcal{Y}$ with $y(0) = 0$ is given by

$$(3.4) \quad \begin{aligned} \mathcal{L}_y(\bar{y}, \bar{\mu}, \bar{p})y &= \int_0^T \int_\Omega \alpha_Q(\bar{y} - y_Q) y \, d\mathbf{x} dt + \int_\Omega \alpha_\Omega(\bar{y}(T) - y_\Omega) y(T) \, d\mathbf{x} \\ &+ \int_0^T \left(\langle y_t, \bar{p} \rangle_{V', V} + \int_\Omega \nabla y \cdot \nabla \bar{p} + d_y(\cdot, \bar{y}(\cdot), \bar{\mu})y\bar{p} \, d\mathbf{x} \right) dt. \end{aligned}$$

From (3.3a) and (3.4) we infer the *adjoint* or *dual equations* for \bar{p} , here written in its strong form:

$$(3.5) \quad \begin{aligned} -\bar{p}_t - \Delta \bar{p} + d_y(\cdot, \bar{y}(\cdot), \bar{\mu})\bar{p} &= -\alpha_Q(\bar{y} - y_Q) && \text{in } Q, \\ \frac{\partial \bar{p}}{\partial n} &= 0 && \text{on } \Sigma, \\ \bar{p}(T) &= \alpha_\Omega(y_\Omega - \bar{y}(T)) && \text{in } \Omega. \end{aligned}$$

Proposition 3.3. *Let Assumption 1 be satisfied. Suppose that $\bar{\mu} \in \mathcal{D}_{ad}$ is a local optimal solution to $(\hat{\mathbf{P}})$ with associated optimal state $\bar{y} = \mathcal{G}(\bar{\mu})$. Then, there exists*

a unique Lagrange multiplier $\bar{p} \in \mathcal{Y}$ satisfying $\bar{p}(T) = \alpha_\Omega(y_\Omega - \bar{y}(T))$ in H and

$$\begin{aligned} & - \langle \bar{p}_t(t), \varphi \rangle_{V', V} + \int_\Omega \nabla \bar{p}(t) \cdot \nabla \varphi + d_y(\cdot, \bar{y}(\cdot), \bar{\mu}) \bar{p}(t) \varphi \, d\mathbf{x} \\ & = \int_\Omega \alpha_Q(t) (y_Q(t) - \bar{y}(t)) \, d\mathbf{x} \quad \text{for all } \varphi \in V \text{ and f.a.a. } t \in [0, T], \end{aligned}$$

where d_y is given in Remark 2.3-2). Moreover, there exists a constant $\hat{C} > 0$ with

$$(3.6) \quad \|\bar{p}\|_{\mathcal{Y}} \leq \hat{C} (\|\alpha_Q(\bar{y} - y_Q)\|_{L^\infty(Q)} + \|\alpha_\Omega(\bar{y}(T) - y_\Omega)\|_{L^\infty(\Omega)}).$$

Proof. For $\bar{\mu} \in \mathcal{D}_{ad}$ and $\bar{y} = \mathcal{G}(\bar{\mu})$ the function

$$d_y(\mathbf{x}, \bar{y}(t, \mathbf{x}), \bar{\mu}) = \sum_{j=1}^m \bar{\mu}_j b_j(\mathbf{x}) \cosh\left(\bar{y}(t, \mathbf{x}) \sum_{i=1}^m \bar{\mu}_i b_i(\mathbf{x})\right), \quad (t, \mathbf{x}) \in Q,$$

is essentially bounded and nonnegative (see Remark 2.3). Thus, the proof follows from Theorem 5.5 in [26, p. 213]. \square

The Fréchet-derivative $\mathcal{L}_\mu(\bar{y}, \bar{\mu}, \bar{p})$ in a direction $\mu \in \mathbb{R}^m$ has the form

$$(3.7) \quad \mathcal{L}_\mu(\bar{y}, \bar{\mu}, \bar{p})\mu = \sum_{i=1}^m \left(\lambda \kappa_i (\bar{\mu}_i - \mu_i^\circ) + \int_0^T \int_\Omega d_{\mu_i}(\cdot, \bar{y}(\cdot), \bar{\mu}) \bar{p} \, d\mathbf{x} dt \right) \mu_i.$$

Combining (3.3b) and (3.7) we derive the *variational inequality*

$$(3.8) \quad \sum_{i=1}^m \left(\lambda \kappa_i (\bar{\mu}_i - \mu_i^\circ) + \int_0^T \int_\Omega d_{\mu_i}(\cdot, \bar{y}(\cdot), \bar{\mu}) \bar{p} \, d\mathbf{x} dt \right) (\mu_i - \bar{\mu}_i) \geq 0$$

for all $\mu = (\mu_1, \dots, \mu_m) \in \mathcal{D}_{ad}$. Summarizing, we infer by standard arguments the following result; see, e.g., [11, 26] for more details.

Theorem 3.4. *Let Assumption 1 hold. Suppose that $\bar{\mu} \in \mathcal{D}_{ad}$ is a local optimal solution to $(\hat{\mathbf{P}})$ with associated optimal state $\bar{y} = \mathcal{G}(\bar{\mu})$. Let \bar{p} denote the associated Lagrange multiplier introduced in Proposition 3.3. Then, first-order necessary optimality conditions for $(\hat{\mathbf{P}})$ are given by the variational inequality (3.8).*

Remark 3.5. Since $\mathcal{L}_y(\bar{y}, \bar{\mu}, \bar{p}) : \mathcal{Y} \rightarrow \mathbb{R}$ is linear and bounded, we write

$$\mathcal{L}_y(\bar{y}, \bar{\mu}, \bar{p})y = \langle \mathcal{L}_y(\bar{y}, \bar{\mu}, \bar{p}), y \rangle_{\mathcal{Y}', \mathcal{Y}} \quad \text{for } y \in \mathcal{Y}.$$

Analogously, $\mathcal{L}_\mu(\bar{y}, \bar{\mu}, \bar{p}) : \mathbb{R}^m \rightarrow \mathbb{R}$ is linear and bounded. Therefore, the derivative $\mathcal{L}_\mu(\bar{y}, \bar{\mu}, \bar{p})$ can be interpreted as a row vector with the components

$$\mathcal{L}_{\mu_i}(\bar{y}, \bar{\mu}, \bar{p}) = \lambda \kappa_i (\bar{\mu}_i - \mu_i^\circ) + \int_0^T \int_\Omega d_{\mu_i}(\cdot, \bar{y}(\cdot), \bar{\mu}) \bar{p} \, d\mathbf{x} dt.$$

Let us define the column vector $\nabla_\mu \mathcal{L}(\bar{y}, \bar{\mu}, \bar{p}) = \mathcal{L}_\mu(\bar{y}, \bar{\mu}, \bar{p})^\top$ as the gradient on \mathcal{L} with respect to μ . \diamond

We can characterize the gradient of the reduced cost functional; see Section 2.4. It follows by standard arguments [11] that the derivative $\hat{J}'(\mu)$ of the reduced cost functional \hat{J} at a given $\mu \in \mathcal{D}_{ad}$ is given by the row vector

$$\hat{J}'(\mu) = (\hat{J}_{\mu_1}(\mu), \dots, \hat{J}_{\mu_m}(\mu))$$

with the components

$$\hat{J}_{\mu_i}(\mu) = \lambda \kappa_i(\mu_i - \mu_i^\circ) + \int_0^T \int_{\Omega} d_{\mu_i}(\cdot, y(\cdot), \mu) p \, d\mathbf{x} dt, \quad i = 1, \dots, m,$$

where $y = \mathcal{G}(\mu)$ holds and p is the weak solution to

$$(3.9) \quad \begin{aligned} -p_t - \Delta p + d_y(\cdot, y(\cdot), \mu) p &= -\alpha_Q(y - y_Q) && \text{in } Q, \\ \frac{\partial p}{\partial n} &= 0 && \text{on } \Sigma, \\ p(T) &= -\alpha_{\Omega}(y(T) - y_{\Omega}) && \text{in } \Omega. \end{aligned}$$

In the sequel we denote by the column vector $\nabla \hat{J}(\mu)^\top$ the gradient of \hat{J} at μ .

3.2. Second-order sufficient optimality conditions. In this section we turn to second-order optimality conditions. For that purpose we make use of the following result, which is proved in the Appendix.

Proposition 3.6. *If Assumption 1-a) holds, the mapping \mathcal{G} is twice continuously Fréchet-differentiable on \mathcal{D}_{ad} . In particular, for $\mu \in \mathcal{D}_{ad}$ the function*

$$z = \mathcal{G}''(\mu)(\mu^1, \mu^2), \quad \mu^1, \mu^2 \in \mathbb{R}^m,$$

satisfies the linear parabolic problem

$$\begin{aligned} z_t - \Delta z + d_y(\cdot, y(\cdot), \mu) z &= -d_{yy}(\cdot, y(\cdot), \mu) y^1 y^2 && \text{in } Q, \\ \frac{\partial z}{\partial n} &= 0 && \text{on } \Sigma, \\ z(0) &= 0 && \text{in } \Omega, \end{aligned}$$

where d_y, d_{yy} are given in Remark 2.3-2) and the directions $y^i, i \in \{1, 2\}$, are the weak solutions to the linear parabolic problems

$$\begin{aligned} y_t^i - \Delta y^i + d_y(\cdot, y(\cdot), \mu) y^i &= -d_{\mu}(\cdot, y(\cdot), \mu) \mu^i && \text{in } Q, \\ \frac{\partial y^i}{\partial n} &= 0 && \text{on } \Sigma, \\ y^i(0) &= 0 && \text{in } \Omega. \end{aligned}$$

Let $\bar{\mu} \in \mathcal{D}_{ad}$ be a local optimal solution to $(\hat{\mathbf{P}})$. By $\bar{y}, \bar{p} \in \mathcal{Y}$ we denote the associated state and adjoint variable, respectively. For a given $\tau > 0$ the set of active constraints at $\bar{\mu}$ is defined as

$$\mathcal{A}_\tau = \left\{ i \in \{1, \dots, m\} : \left| \lambda \kappa_i(\bar{\mu}_i - \mu_i^\circ) + \int_0^T \int_{\Omega} d_{\mu_i}(\cdot, \bar{y}(\cdot), \bar{\mu}) \bar{p} \, d\mathbf{x} dt \right| > \tau \right\}.$$

Further, the critical cone $\mathcal{C}_\tau(\bar{\mu})$ is the set of all $\mu \in \mathbb{R}^m$ satisfying for $i \in \{1, \dots, m\}$

$$\mu_i \begin{cases} = 0 & \text{if } i \in \mathcal{A}_\tau(\bar{\mu}), \\ \geq 0 & \text{if } \bar{\mu}_i = \mu_i^a \text{ and } i \notin \mathcal{A}_\tau(\bar{\mu}), \\ \leq 0 & \text{if } \bar{\mu}_i = \mu_i^b \text{ and } i \notin \mathcal{A}_\tau(\bar{\mu}). \end{cases}$$

Notice that the Lagrangian is twice continuously Fréchet-differentiable. By \mathcal{L}_{xx} we denote the second derivative with respect to the pair $x = (y, \mu)$. Then, a second-order sufficient optimality condition is formulated in the next theorem.

Theorem 3.7. *Let Assumption 1 hold. Suppose that $\bar{\mu} \in \mathcal{D}_{ad}$ is a local solution to $(\hat{\mathbf{P}})$ and $\bar{y} = \mathfrak{G}(\bar{\mu})$ denotes the associated optimal state. Let $\bar{p} \in \mathcal{Y}$ be the corresponding Lagrange multiplier introduced in Proposition 3.3. Assume that there exist scalars $\tau > 0$ and $\delta > 0$ satisfying the second-order condition*

$$(3.10) \quad \mathcal{L}_{xx}(\bar{y}, \bar{\mu}, \bar{p})(x, x) \geq \delta |\mu|_2^2 \quad \text{for all } x = (y, \mu) \in W(0, T) \times \mathcal{C}_\tau(\bar{\mu})$$

with

$$\begin{aligned} y_t - \Delta y + d_y(\cdot, \bar{y}(\cdot), \bar{\mu})y &= -d_\mu(\cdot, \bar{y}(\cdot), \bar{\mu})\mu && \text{in } Q, \\ \frac{\partial y}{\partial n} &= 0 && \text{on } \Sigma, \\ y(0) &= 0 && \text{in } \Omega. \end{aligned}$$

Then, there are scalars $\varepsilon > 0$, $\sigma > 0$ such that

$$\hat{J}(\mu) \geq \hat{J}(\bar{\mu}) + \sigma |\mu - \bar{\mu}|_2^2 \quad \text{for all } \mu \in \mathcal{D}_{ad} \text{ with } |\mu - \bar{\mu}|_2 \leq \varepsilon.$$

Hence, $\bar{\mu}$ is a strict local minimizer.

For a proof of Theorem 3.7 we refer the reader to [26, Theorem 5.17]. To give sufficient conditions for (3.10) we derive the second derivative of the Lagrangian:

$$\begin{aligned} \mathcal{L}_{yy}(\bar{y}, \bar{\mu}, \bar{p})(y, \tilde{y}) &= \int_0^T \int_\Omega (\alpha_Q + d_{yy}(\cdot, \bar{y}(\cdot), \bar{\mu})\bar{p})y\tilde{y} \, d\mathbf{x}dt + \int_\Omega \alpha_\Omega y(T)\tilde{y}(T) \, d\mathbf{x}, \\ \mathcal{L}_{\mu\mu}(\bar{y}, \bar{\mu}, \bar{p})(\mu, \tilde{\mu}) &= \tilde{\mu}^\top \left(\lambda \text{diag}(\kappa_1, \dots, \kappa_m) + \int_0^T \int_\Omega d_{\mu\mu}(\cdot, \bar{y}(\cdot), \bar{\mu})\bar{p} \, d\mathbf{x}dt \right) \mu, \\ \mathcal{L}_{y\mu}(\bar{y}, \bar{\mu}, \bar{p})(y, \mu) &= \int_0^T \int_\Omega (d_{y\mu}(\cdot, \bar{y}(\cdot), \bar{\mu})y)\mu\bar{p} \, d\mathbf{x}dt = \mathcal{L}_{\mu y}(\bar{y}, \bar{\mu}, \bar{p})(\mu, y), \end{aligned}$$

where the symmetric Hessian matrix $d_{\mu\mu}(\cdot, \bar{y}(\cdot), \bar{\mu})$ is an $m \times m$ matrix given by

$$(d_{\mu\mu}(\cdot, \bar{y}(\cdot), \bar{\mu}))_{ij} = \sinh\left(\bar{y}(\cdot) \sum_{l=1}^m \bar{\mu}_l b_l\right) \bar{y}^2(\cdot) b_i b_j, \quad 1 \leq i, j \leq m$$

and the mixed derivative reads

$$\begin{aligned} (d_{y\mu}(\cdot, \bar{y}(\cdot), \bar{\mu})y)\mu &= \sum_{i=1}^m (d_{y\mu_i}(\cdot, \bar{y}(\cdot), \bar{\mu})y)\mu_i \\ &= \sum_{i=1}^m \left(\left(\bar{y}(\cdot) \sum_{l=1}^m \bar{\mu}_l b_l \sinh\left(\bar{y}(\cdot) \sum_{j=1}^m \bar{\mu}_j b_j\right) + \cosh\left(\bar{y}(\cdot) \sum_{j=1}^m \bar{\mu}_j b_j\right) \right) b_i y(\cdot) \right) \mu_i. \end{aligned}$$

We set $\eta = \lambda \min_{1 \leq i \leq m} \kappa_i$ and suppose that $\kappa > 0$ holds. Using $(\bar{y}, \bar{\mu}) \in L^\infty(Q) \times \mathcal{D}_{ad}$ we define

$$\begin{aligned} C_1 &= |Q| \|d_{yy}(\cdot, \bar{y}(\cdot), \bar{\mu})\|_{L^\infty(Q)} < \infty, \\ C_2 &= 2|Q| \sqrt{\sum_{i=1}^m \|d_{y\mu_i}(\cdot, \bar{y}(\cdot), \bar{\mu})\|_{L^\infty(Q)}^2} < \infty, \\ C_3 &= |Q| \max_Q \|d_{\mu\mu}(\cdot, \bar{y}(\cdot), \bar{\mu})\|_2 < \infty, \end{aligned}$$

where $|Q|$ stands for the finite Lebesgue measure of Q and $\|\cdot\|_2$ denotes the spectral norm for symmetric matrices. Recall that α_Q and α_Ω are nonnegative. Moreover,

$$\int_0^T \int_\Omega \varphi \psi \, d\mathbf{x} dt \leq \|\varphi\|_{L^2(Q)} \|\psi\|_{L^2(Q)} \leq |Q| \|\varphi\|_{L^\infty(Q)} \|\psi\|_{L^\infty(Q)} \leq |Q| \|\varphi\|_{\mathcal{Y}} \|\psi\|_{\mathcal{Y}}$$

for $\varphi, \psi \in \mathcal{Y}$. Thus, applying (3.2) and (3.6) we infer that

$$\begin{aligned} \mathcal{L}_{xx}(\bar{y}, \bar{\mu}, \bar{p})(x, x) &= \mathcal{L}_{yy}(\bar{y}, \bar{\mu}, \bar{p})(y, y) + 2\mathcal{L}_{y\mu}(\bar{y}, \bar{\mu}, \bar{p})(y, \mu) + \mathcal{L}_{\mu\mu}(\bar{y}, \bar{\mu}, \bar{p})(\mu, \mu) \\ &\geq \eta |\mu|_2^2 + \int_0^T \int_\Omega (d_{yy}(\cdot, \bar{y}(\cdot), \bar{\mu}) y^2 + 2(d_{y\mu}(\cdot, \bar{y}(\cdot), \bar{\mu}) y) \mu) \bar{p} \, d\mathbf{x} dt \\ &\quad + \int_0^T \int_\Omega (\mu^\top d_{\mu\mu}(\cdot, \bar{y}(\cdot), \bar{\mu}) \mu) \bar{p} \, d\mathbf{x} dt \\ &\geq \eta |\mu|_2^2 - C_1 \|y\|_{\mathcal{Y}}^2 \|\bar{p}\|_{\mathcal{Y}} - C_2 |\mu|_2 \|y\|_{\mathcal{Y}} \|\bar{p}\|_{\mathcal{Y}} - C_3 |\mu|_2^2 \|\bar{p}\|_{\mathcal{Y}} \\ &\geq \left(\eta - C_4 (\|\alpha_Q(\bar{y} - y_Q)\|_{L^\infty(Q)} + \|\alpha_\Omega(\bar{y}(T) - y_\Omega)\|_{L^\infty(\Omega)}) \right) |\mu|_2^2 \end{aligned}$$

for all $(y, \mu) \in W(0, T) \times \mathbb{R}^m$ satisfying (3.1), where we put $C_4 = (C_1 \bar{C}^2 + C_2 \bar{C} + C_3) \bar{C}$. Thus, (3.10) holds provided

$$\|\alpha_Q(\bar{y} - y_Q)\|_{L^\infty(Q)} + \|\alpha_\Omega(\bar{y}(T) - y_\Omega)\|_{L^\infty(\Omega)} < \frac{\eta}{2C_4}.$$

Summarizing, we have shown the following result.

Proposition 3.8. *Let Assumption 1 hold. Further, the regularization parameters $\lambda, \kappa_1, \dots, \kappa_m$ are positive. Suppose that $\bar{\mu} \in \mathcal{D}_{ad}$ is a local solution to $(\hat{\mathbf{P}})$ and $\bar{y} = \mathcal{G}(\bar{\mu})$ denotes the associated optimal state. Let $\bar{p} \in \mathcal{Y}$ be the corresponding Lagrange multiplier introduced in Proposition 3.3. If the residuum $\|\alpha_Q(\bar{y} - y_Q)\|_{L^\infty(Q)} + \|\alpha_\Omega(\bar{y}(T) - y_\Omega)\|_{L^\infty(\Omega)}$ is sufficiently small, (3.10) is satisfied, i.e., $\bar{\mu}$ is a strict local minimizer for $(\hat{\mathbf{P}})$.*

3.3. Representation of the Hessian. Next we derive an expression for the Hessian $\hat{J}''(\mu) \in \mathbb{R}^{m \times m}$ for an arbitrary $\mu \in \mathcal{D}_{ad}$; see [11], for instance. We have the identity

$$\hat{J}(\mu) = J(\mathcal{G}(\mu), \mu) = \mathcal{L}(\mathcal{G}(\mu), \mu, p) \quad \text{for } p \in L^2(0, T; V).$$

Differentiating \hat{J} in a direction $\mu^1 \in \mathbb{R}^m$ yields

$$\begin{aligned} \langle \nabla \hat{J}(\mu), \mu^1 \rangle_{\mathbb{R}^m} &= \hat{J}'(\mu) \mu^1 = \mathcal{L}_y(\mathcal{G}(\mu), \mu, p) \mathcal{G}'(\mu) \mu^1 + \mathcal{L}_\mu(\mathcal{G}(\mu), \mu, p) \mu^1 \\ &= \langle \mathcal{L}_y(\mathcal{G}(\mu), \mu, p), \mathcal{G}'(\mu) \mu^1 \rangle_{\mathcal{Y}', \mathcal{Y}} + \langle \nabla_\mu \mathcal{L}(\mathcal{G}(\mu), \mu, p) \mu^1 \rangle_{\mathbb{R}^m}. \end{aligned}$$

For $\mu^1, \mu^2 \in \mathbb{R}^m$ we find:

$$\begin{aligned} \langle \hat{J}''(\mu) \mu^1, \mu^2 \rangle_{\mathbb{R}^m} &= \langle \mathcal{L}_y(\mathcal{G}(\mu), \mu, p), \mathcal{G}''(\mu)(\mu^1, \mu^2) \rangle_{\mathcal{Y}', \mathcal{Y}} \\ &\quad + \langle \mathcal{L}_{yy}(\mathcal{G}(\mu), \mu, p) \mathcal{G}'(\mu) \mu^1, \mathcal{G}'(\mu) \mu^2 \rangle_{\mathcal{Y}', \mathcal{Y}} \\ &\quad + \langle \mathcal{L}_{y\mu}(\mathcal{G}(\mu), \mu, p) \mu^1, \mathcal{G}'(\mu) \mu^2 \rangle_{\mathcal{Y}', \mathcal{Y}} \\ &\quad + \langle \mathcal{L}_{\mu y}(\mathcal{G}(\mu), \mu, p) \mathcal{G}'(\mu) \mu^1, \mu^2 \rangle_{\mathbb{R}^m} \\ &\quad + \langle \mathcal{L}_{\mu\mu}(\mathcal{G}(\mu), \mu, p) \mu^1, \mu^2 \rangle_{\mathbb{R}^m}. \end{aligned}$$

Now we choose for $p \in L^2(0, T, V)$ the solution $p(\mu)$ to the adjoint equation (3.5), i.e., for $p = p(\mu)$ holds

$$\mathcal{L}_y(\mathcal{G}(\mu), \mu, p(\mu)) = 0 \quad \text{in } \mathcal{Y}'.$$

Hence, the term containing $\mathcal{G}''(\mu)$ drops out and by rearranging the dual pairings we obtain

$$\begin{aligned} & \langle \hat{J}''(\mu)\mu^1, \mu^2 \rangle_{\mathbb{R}^m} \\ &= \langle (\mathcal{G}'(\mu)^* \mathcal{L}_{yy}(\mathcal{G}(\mu), \mu, p(\mu)) \mathcal{G}'(\mu) + \mathcal{G}'(\mu)^* \mathcal{L}_{y\mu}(\mathcal{G}(\mu), \mu, p(\mu)))\mu^1, \mu^2 \rangle_{\mathbb{R}^m} \\ & \quad + \langle (\mathcal{L}_{\mu y}(\mathcal{G}(\mu), \mu, p(\mu)) \mathcal{G}'(\mu) + \mathcal{L}_{\mu\mu}(\mathcal{G}(\mu), \mu, p(\mu)))\mu^1, \mu^2 \rangle_{\mathbb{R}^m}, \end{aligned}$$

where $\mathcal{G}'(\mu)^* : \mathcal{Y}' \rightarrow (\mathbb{R}^m)' \simeq \mathbb{R}^m$ denotes the dual operator of $\mathcal{G}(\mu)$ satisfying

$$\langle \mathcal{G}'(\mu)^* r, \mu \rangle_{\mathbb{R}^m} = \langle r, \mathcal{G}'(\mu)\mu \rangle_{\mathcal{Y}', \mathcal{Y}} \quad \text{for all } r \in \mathcal{Y}' \text{ and } \mu \in \mathbb{R}^m.$$

Consequently, the second derivative of the reduced cost function \hat{J} as follows:

$$(3.11) \quad \begin{aligned} \hat{J}''(\mu) &= \mathcal{G}'(\mu)^* \mathcal{L}_{yy}(y(\mu), \mu, p(\mu)) \mathcal{G}'(\mu) + \mathcal{G}'(\mu)^* \mathcal{L}_{y\mu}(y(\mu), \mu, p(\mu)) \\ & \quad + \mathcal{L}_{\mu y}(y(\mu), \mu, p(\mu)) \mathcal{G}'(\mu) + \mathcal{L}_{\mu\mu}(y(\mu), \mu, p(\mu)). \end{aligned}$$

This can be formulated in the following way:

$$(3.12) \quad \hat{J}''(\mu) = T(\mu)^* \mathcal{L}_{xx}(\mathcal{G}(\mu), \mu, p(\mu)) T(\mu), \quad x = (y, \mu),$$

with the operator

$$T(\mu) = \begin{pmatrix} \mathcal{G}'(\mu) \\ I \end{pmatrix} \in L(\mathbb{R}^m, \mathcal{Y} \times \mathbb{R}^m),$$

the dual operator

$$T(\mu)^* = \begin{pmatrix} \mathcal{G}'(\mu) \\ I \end{pmatrix}^* = (\mathcal{G}'(\mu)^* \quad I) \in L(\mathcal{Y}' \times \mathbb{R}^m, \mathbb{R}^m),$$

and the second derivative

$$\mathcal{L}_{xx} = \begin{pmatrix} \mathcal{L}_{yy} & \mathcal{L}_{y\mu} \\ \mathcal{L}_{\mu y} & \mathcal{L}_{\mu\mu} \end{pmatrix} \in L(\mathcal{Y} \times \mathbb{R}^m, \mathcal{Y}' \times \mathbb{R}^m).$$

Here, $L(\mathbb{R}^m, \mathcal{Y} \times \mathbb{R}^m)$ denotes the Banach space of all linear and bounded operators from $\mathcal{Y}' \times \mathbb{R}^m$ to \mathbb{R}^m endowed with the usual operator norm, the mapping $I \in L(\mathbb{R}^m)$ is the identity in \mathbb{R}^m . Throughout the paper we utilize the notation

$$\hat{J}''(\mu)(\mu^1, \mu^2) = \langle \hat{J}''(\mu)\mu^1, \mu^2 \rangle_{\mathbb{R}^m} = (\mu^2)^\top \hat{J}''(\mu)\mu^1, \quad \mu^1, \mu^2 \in \mathbb{R}^m,$$

for the Hessian of the reduced cost functional.

Remark 3.9. By this approach we do not use the Hessian representation (3.12) to set up the Hessian matrix explicitly. In fact we compute just the “effect” of the operator $\hat{J}''(\hat{\mu})$ at $\hat{\mu} \in \mathcal{D}_{ad}$ on a direction $\mu \in \mathbb{R}^m$ by applying the operator components of (3.12) consecutively. Therefore, for given $\mu \in \mathbb{R}^m$, we have to solve the *linearized state equations*

$$(3.13) \quad \begin{aligned} y_t - \Delta y + \left(\sum_{i=1}^m \hat{\mu}_i b_i \cosh \left(\hat{y} \sum_{j=1}^m \hat{\mu}_j b_j \right) \right) y &= \hat{y} \cosh \left(\hat{y} \sum_{j=1}^m \hat{\mu}_j b_j \right) \sum_{i=1}^m \mu_i b_i & \text{in } Q \\ \frac{\partial y}{\partial n} &= 0 & \text{on } \Sigma \\ y(0) &= 0 & \text{in } \Omega \end{aligned}$$

with $\hat{y} = y(\hat{\mu}) = \mathcal{G}(\hat{\mu})$. Let \hat{p} be the weak solution to (3.9) for $y = \hat{y}$ and $\mu = \hat{\mu}$. We set $\hat{x} = (\hat{y}, \hat{\mu})$. In the next step, we compute

$$\mathcal{L}_{xx}(\hat{x}, \hat{p})(y, \mu) = \begin{pmatrix} \mathcal{L}_{yy}(\hat{x}, \hat{p})y + \mathcal{L}_{y\mu}(\hat{x}, \hat{p})\mu \\ \mathcal{L}_{\mu y}(\hat{x}, \hat{p})y + \mathcal{L}_{\mu\mu}(\hat{x}, \hat{p})\mu \end{pmatrix} = \begin{pmatrix} h_1 \\ h_2 \end{pmatrix} \in \mathcal{Y}' \times \mathbb{R}^m,$$

and apply the operator $T(\hat{\mu})^*$ on the result, i.e.

$$T(\hat{\mu})^* \begin{pmatrix} h_1 \\ h_2 \end{pmatrix} = (y'(\hat{\mu})^*, I_{\mathbb{R}^m}) \begin{pmatrix} h_1 \\ h_2 \end{pmatrix} = h_3 + h_4.$$

While for the second component h_4 of the result we obtain immediately $h_4 = h_2$ (since I_U is the identity in \mathbb{R}^m), the situation for the first component h_3 is a bit more complicated: We introduce the temporary variable $\tilde{h}_3 = (\tilde{h}_3^1, \tilde{h}_3^2) \in L^2(0, T; V) \times H$ and solve the *adjoint equations*

$$\begin{aligned} -(\tilde{h}_3^1)_t - \Delta \tilde{h}_3^1 + \left(\cosh \left(\hat{y} \sum_{i=1}^m \hat{\mu}_i b_i \right) \sum_{i=1}^m \hat{\mu}_i b_i \right) \tilde{h}_3^1 \\ = -\alpha_Q y - \sinh \left(\hat{y} \sum_{i=1}^m \hat{\mu}_i b_i \right) \left(y \sum_{i=1}^m \hat{\mu}_i b_i \right) \left(\hat{p} \sum_{i=1}^m \hat{\mu}_i b_i \right) \\ - \sinh \left(\hat{y} \sum_{i=1}^m \hat{\mu}_i b_i \right) \left(\hat{y} \sum_{i=1}^m \mu_i b_i \right) \left(\hat{p} \sum_{i=1}^m \hat{\mu}_i b_i \right) \\ - \cosh \left(\hat{y} \sum_{i=1}^m \hat{\mu}_i b_i \right) \left(\hat{y} \sum_{i=1}^m \mu_i b_i \right) \left(\hat{p} \sum_{i=1}^m \hat{\mu}_i b_i \right) \quad \text{in } Q \\ \frac{\partial \tilde{h}_3^1}{\partial n} = 0 \quad \text{on } \Sigma \\ \tilde{h}_3^1(T) = -\alpha_\Omega y(T) \quad \text{in } \Omega \end{aligned}$$

and set $\tilde{h}_3^2 = \tilde{h}_3^1(0)$. Thus, we can compute the entries of h_3 in the following way:

$$h_{3i} = \int_0^T \int_\Omega \cosh \left(y(t) \sum_{j=1}^m \mu_j b_j \right) (y(t) b_i) \tilde{h}_3^1(t) \, d\mathbf{x} dt \quad \text{for } i = 1, \dots, m.$$

The entries of h_4 are given as

$$\begin{aligned} h_{4j} = \int_0^T \int_\Omega \left(y \hat{y} b_j \sum_{i=1}^m \hat{\mu}_i b_i \sinh \left(\hat{y} \sum_{i=1}^m \hat{\mu}_i b_i \right) + y b_j \cosh \left(\hat{y} \sum_{i=1}^m \hat{\mu}_i b_i \right) \right) \hat{p} \, d\mathbf{x} dt \\ + \lambda \kappa_j \mu_j + \int_0^T \int_\Omega \hat{y}^2 b_i \sum_{i=1}^m \hat{\mu}_i b_i \sinh \left(\hat{y} \sum_{i=1}^m \hat{\mu}_i b_i \right) \hat{p} \, d\mathbf{x} dt \quad \text{for } j = 1, \dots, m. \end{aligned}$$

Finally as a result we obtain $\hat{J}''(\hat{\mu})\mu = h_3 + h_4$. \diamond

4. A-POSTERIORI ERROR ESTIMATES

In this section we want to introduce the main idea underlying our a-posteriori error analysis for nonlinear optimal control problems: supposing that $\tilde{\mu}$ is an arbitrary element of the admissible parameter set \mathcal{D}_{ad} , our aim is now to estimate the difference $|\bar{\mu} - \tilde{\mu}|_2$ without knowing the optimal solution $\bar{\mu}$. The associated idea is not new and was used, for instance, in the context of error estimates for the optimal control of ODEs by Malanowski et al. [21]. An a-posteriori error estimate for linear-quadratic optimal control problems of PDEs under application of proper orthogonal decomposition as a model order reduction technique was investigated in [27] and extended to some nonlinear case in [15]. We briefly recall the basic idea:

If $\tilde{\mu} \neq \bar{\mu}$, i.e. $\tilde{\mu}$ is not an optimal solution, then $\tilde{\mu}$ does not satisfy the necessary optimality condition (3.8). Nevertheless, there exists a function $\zeta = (\zeta_1, \dots, \zeta_m) \in$

\mathbb{R}^m such that

$$(4.1) \quad \sum_{i=1}^m \left(\lambda \kappa_i (\tilde{\mu}_i - \mu_i^\circ) + \int_0^T \int_{\Omega} \cosh \left(\tilde{y} \sum_{j=1}^m \tilde{\mu}_j b_j \right) b_i \tilde{y} \tilde{p} \, d\mathbf{x} dt + \zeta_i \right) (\mu_i - \tilde{\mu}_i) \geq 0$$

is fulfilled for all $\mu \in \mathcal{D}_{ad}$, i.e., $\tilde{\mu}$ satisfies the optimality condition of a ‘‘perturbed’’ semilinear parabolic optimal control problem with perturbation ζ . In (4.1) we have $\tilde{y} = \mathfrak{G}(\tilde{\mu})$ and \tilde{p} solve the adjoint equation for the parameter $\tilde{\mu}$ and the associated state \tilde{y} . The smaller ζ is, the closer is $\tilde{\mu}$ to the optimal parameter $\bar{\mu}$.

An estimation for the distance of $|\bar{\mu} - \tilde{\mu}|_2$ in terms of the perturbation ζ for linear-quadratic optimal control problems is achieved in [27, Theorem 3.1], while an estimation for nonlinear problems is derived in [15, Theorem 2.5]. For the latter case, the situation is more complicated and one has to put more effort to determine a suitable estimate. This is due to the fact that some second-order information on $\bar{\mu}$ is needed.

Assume that there exists some constant $\delta > 0$ such that the coercivity condition

$$(4.2) \quad \hat{J}''(\bar{\mu})(\mu, \mu) \geq \delta |\mu|_2^2 \quad \text{for all } \mu \in \mathbb{R}^m$$

is satisfied. Then for any $0 < \delta' < \delta$ there exists a radius $\rho(\delta') > 0$ such that

$$(4.3) \quad \hat{J}''(\bar{\mu})(\mu, \mu) \geq \delta' |\mu|_2^2 \quad \text{for all } \mu \text{ with } |\mu - \bar{\mu}|_2 \leq \rho(\delta'), \text{ for all } v \in \mathbb{R}^m.$$

Since we are interested in the order of the error, we follow the proposal in [15, Remark 2.3] and select $\delta' := \delta/2$ and set $\rho := \rho(\delta/2)$. If $\tilde{\mu}$ belongs to this neighborhood, we can estimate the distance in the following way:

$$(4.4) \quad |\bar{\mu} - \tilde{\mu}|_2 \leq \frac{2}{\delta} |\zeta|_2.$$

For a proof we refer to [15].

We proceed by constructing the function ζ . Suppose that we have $\tilde{\mu}$ and the associated adjoint state \tilde{p} . The goal is to determine $\zeta \in \mathbb{R}^m$ satisfying the perturbed variational inequality (4.1). This is fulfilled by defining ζ in the following way

$$(4.5) \quad \zeta_i := \begin{cases} \left[\lambda \kappa_i (\tilde{\mu}_i - \mu_i^\circ) + \int_0^T \int_{\Omega} d_{\mu_i}(\cdot, \tilde{y}(\cdot), \mu) \tilde{p} \, d\mathbf{x} dt \right]_-, & \text{if } \tilde{\mu}_i = \mu_i^a, \\ - \left[\lambda \kappa_i (\tilde{\mu}_i - \mu_i^\circ) + \int_0^T \int_{\Omega} d_{\mu_i}(\cdot, \tilde{y}(\cdot), \mu) \tilde{p} \, d\mathbf{x} dt \right], & \text{if } \mu_i^a < \tilde{\mu}_i < \mu_i^b, \\ - \left[\lambda \kappa_i (\tilde{\mu}_i - \mu_i^\circ) + \int_0^T \int_{\Omega} d_{\mu_i}(\cdot, \tilde{y}(\cdot), \mu) \tilde{p} \, d\mathbf{x} dt \right]_+, & \text{if } \tilde{\mu}_i = \mu_i^b \end{cases}$$

for $i = 1, \dots, m$. Here $[s]_- = -\min(0, s)$ denotes the negative part function and $[s]_+ = \max(0, s)$ denotes the positive part function.

Next, we need an approximation of the coercivity constant δ . For this reason, the approximation of the Hessian \hat{J}'' associated with the suboptimal parameter $\tilde{\mu}$ is taken into account. We have to assume that $\hat{J}''(\tilde{\mu})$ is positive definite. Let σ_{\min} be the smallest eigenvalue of $\hat{J}''(\tilde{\mu})$. Then there holds

$$\mu^\top \hat{J}''(\tilde{\mu}) \mu \geq \sigma_{\min} |\mu|_2^2 \quad \text{for all } \mu \in \mathbb{R}^m.$$

Hence, if the control problem behaves well around $\bar{\mu}$, the coercivity constant δ can be approximated by σ_{\min} . Assuming that

$$\sigma_{\min} \leq \delta$$

holds, we can deduce that the distance of $\tilde{\mu}$ to the unknown locally optimal parameter $\bar{\mu}$ can be estimated by

$$(4.6) \quad |\tilde{\mu} - \bar{\mu}|_2 \leq \frac{2}{\sigma_{\min}} |\zeta|_2.$$

We will call (4.6) an a-posteriori error estimate, since, in the next section, we shall apply it to suboptimal solutions $\tilde{\mu}$ that have already been computed from a POD model. After having computed \tilde{u} , we determine the associated state \tilde{y} and the Lagrange multiplier \tilde{p} . Then, we can determine ζ as well as its Euclidean norm and (4.6) gives an upper bound for the distance of $\tilde{\mu}$ to $\bar{\mu}$. In this way, the error caused by the POD approximation can be estimated a-posteriorily. If the error is too large, then we have to improve the POD basis, e.g., by including more POD basis functions in our Galerkin ansatz.

5. THE POD GALERKIN METHOD

Let $\mu \in \mathcal{D}_{ad}$ be chosen arbitrarily and $y = \mathcal{G}(\mu)$. We denote by X either the Hilbert space V or H . For $\ell \in \mathbb{N}$ we consider the minimization problem

$$(5.1) \quad \begin{aligned} \min_{\tilde{\psi}_1, \dots, \tilde{\psi}_\ell \in X} \int_0^T \left\| y(t) - \sum_{i=1}^{\ell} \langle y(t), \tilde{\psi}_i \rangle_X \tilde{\psi}_i \right\|_X^2 dt \\ \text{s.t. } \langle \tilde{\psi}_i, \tilde{\psi}_j \rangle_X = \delta_{ij} \text{ for } 1 \leq i, j \leq \ell. \end{aligned}$$

A solution $\{\psi_i\}_{i=1}^{\ell}$ to (5.1) is called *POD basis of rank ℓ* . We introduce the integral operator $\mathcal{R} : X \rightarrow X$ as

$$\mathcal{R}\psi = \int_0^T \langle y(t), \psi \rangle_X y(t) dt \quad \text{for } \psi \in X.$$

which is a linear, compact, self-adjoint and nonnegative operator; see, e.g., [13, 10]. Hence, there exists a complete set $\{\psi_i\}_{i=1}^{\infty} \subset X$ of eigenfunctions and associated eigenvalues $\{\lambda_i\}_{i=1}^{\infty}$ satisfying

$$\mathcal{R}\psi_i = \lambda_i \psi_i \text{ for } i = 1, 2, \dots \quad \text{and} \quad \lambda_1 \geq \lambda_2 \geq \dots \geq 0 \text{ with } \lim_{i \rightarrow \infty} \lambda_i = 0$$

It is proved in [13, 10], for instance, that the first ℓ eigenfunctions $\{\psi_i\}_{i=1}^{\ell}$ solve (5.1) and

$$(5.2) \quad \int_0^T \left\| y(t) - \sum_{i=1}^{\ell} \langle y(t), \psi_i \rangle_X \psi_i \right\|_X^2 dt = \sum_{i=\ell+1}^{\infty} \lambda_i$$

holds.

Suppose that for given $\mu \in \mathcal{D}_{ad}$ we have determined a POD basis $\{\psi_i\}_{i=1}^{\ell}$ of rank ℓ . We define the X -orthogonal projection operator

$$\mathcal{P}^{\ell} \varphi = \sum_{i=1}^{\ell} \langle \varphi, \psi_i \rangle_X \psi_i \quad \text{for } \varphi \in X.$$

Then a POD Galerkin scheme for (2.1b)-(2.1d) is given as follows: $y^{\ell}(t) = \sum_{j=1}^{\ell} y_j^{\ell}(t) \psi_j$, $t \in [0, T]$ a.e., solves

$$(5.3a) \quad \langle y_t^{\ell}(t), \psi_i \rangle_{V', V} + \int_{\Omega} \nabla y^{\ell}(t) \cdot \nabla \psi_i + d(\cdot, y^{\ell}(t), \mu) \psi_i \, d\mathbf{x} = \int_{\Omega} f(t) \psi_i \, d\mathbf{x}$$

for $1 \leq i \leq \ell$ and f.a.a. $t \in [0, T]$ and

$$(5.3b) \quad y^\ell(0) = \mathcal{P}^\ell y_o.$$

Remark 5.1. The numerical evaluation of the nonlinear terms

$$\int_{\Omega} d(\cdot, y^\ell(t), \mu) \psi_i \, d\mathbf{x}, \quad 1 \leq i \leq \ell,$$

is expensive, so that we apply the empirical interpolation method (EIM) [9, 22] in our numerical experiments. For an easier presentation, we do the POD Galerkin scheme without EIM. We also refer to [19] for more details. \diamond

For a proof of the next proposition we refer the reader to the Appendix.

Proposition 5.2. *Let Assumption 1-a) be satisfied. Then (5.3) has a unique solution for every $\mu \in \mathcal{D}_{ad}$.*

Next we present an a-priori error estimate for the POD Galerkin scheme which is proved in the Appendix.

Proposition 5.3. *Suppose that Assumption 1-a) holds. For $\mu \in \mathcal{D}_{ad}$ let y and y^ℓ are the solutions to (2.3) and (5.3), respectively. Then there exists a constant such that*

$$(5.4) \quad \int_0^T \|y(t) - y^\ell(t)\|_X^2 \, dt \leq C \left(\sum_{i=\ell+1}^{\infty} \lambda_i + \|\mathcal{P}^\ell y_t(t) - y_t(t)\|_{V'}^2 \right),$$

where we take the Hilbert space $X = V$ in (5.1).

Analogously to the state equation we determine a POD basis for the adjoint equation. Let $\mu \in \mathcal{D}_{ad}$ and $y = \mathcal{G}(\mu)$. Suppose that p is the weak solution to the adjoint equation

$$\begin{aligned} -p_t - \Delta p + d_y(\cdot, y(\cdot), \mu)p &= -\alpha_Q(y - y_Q) && \text{in } Q, \\ \frac{\partial p}{\partial n} &= 0 && \text{on } \Sigma, \\ p(T) &= -\alpha_\Omega(y(T) - y_\Omega) && \text{in } \Omega. \end{aligned}$$

Then, for $\varphi \in \mathbb{N}$ the POD basis $\{\phi_i\}_{i=1}^\varphi$ of rank φ for the adjoint variable is the solution to

$$\begin{aligned} \min_{\tilde{\phi}_1, \dots, \tilde{\phi}_\varphi \in X} \int_0^T \left\| p(t) - \sum_{i=1}^\varphi \langle p(t), \tilde{\phi}_i \rangle_X \tilde{\phi}_i \right\|_X^2 \, dt \\ \text{s.t. } \langle \tilde{\phi}_i, \tilde{\phi}_j \rangle_X = \delta_{ij} \text{ for } 1 \leq i, j \leq \varphi. \end{aligned}$$

Remark 5.4. The POD Galerkin scheme for the adjoint equation is introduced in a similar manner than for the state equation. From the property $d_y(\mathbf{x}, y, \mu) \geq 0$ f.a.a. $\mathbf{x} \in \Omega$ and for all $(y, \mu) \in \mathbb{R} \times \mathcal{D}_{ad}$ we can derive a-priori bounds for the solution p^ℓ to the POD Galerkin scheme for the state equation. Furthermore, the $L^2(0, T; X)$ -norm of the difference $p - p^\ell$ is bounded by the sum over the eigenvalues of the neglected eigenfunctions as well as by norms of the differences $y - y^\ell$ and $\|p_t - \mathcal{P}^\varphi p_t\|$, where $\mathcal{P}^\varphi \varphi = \sum_{i=1}^\varphi \langle \varphi, \phi_i \rangle_V \phi_i$, $\varphi \in V$, is the orthogonal projection from V onto the finite-dimensional subspace $V^\varphi = \text{span}\{\phi_1, \dots, \phi_\varphi\}$. For more details we refer the reader to [12, 27]. \diamond

6. IMPLEMENTATIONS

In this section we state the algorithms for solving the optimization problem (2.1) and for computing the a-posteriori error estimator.

6.1. The Adaptive POD-OPT algorithm. For solving the optimization problem (2.1) we implemented an adaptive optimization algorithm using POD (Algorithm 1).

Algorithm 1 Adaptive POD-OPT algorithm (Adaptive optimization algorithm using POD)

Input: $\mu^0, \varepsilon, \ell, \varepsilon^{POD}, \sigma$

Output: μ, y

```

1:  $k \leftarrow 0$ 
2:  $y^k \leftarrow$  solve (state equation) for  $\mu^k$  using FEM
3:  $p^k \leftarrow$  solve (adjoint equation) for  $\mu^k$  and  $y^k$  using FEM
4:  $\{\psi_j\}_{j=1}^\ell \leftarrow$  compute POD basis from snapshots  $[y^k, p^k]$ 
5:  $\nabla J^k \leftarrow$  evaluate (reduced gradient) for  $\mu^k$ 
6: while  $\|\nabla J^k\| > \varepsilon$  do
7:    $d^k \leftarrow$  compute search direction using a Newton-CG method
8:    $\mu^{k+1} \leftarrow \mu^k + d^k$ 
9:    $y^{k+1} \leftarrow$  solve (state equation) for  $\mu^{k+1}$  using ROM
10:   $\rho \leftarrow$  evaluate error indicator for  $y^{k+1}$ 
11:  if  $\rho > \varepsilon^{POD}$  then
12:     $y^{k+1} \leftarrow$  solve (state equation) for  $\mu^{k+1}$  using FEM
13:     $p^{k+1} \leftarrow$  solve (adjoint equation) for  $\mu^{k+1}$  and  $y^{k+1}$  using FEM
14:     $\{\psi_j\}_{j=1}^\ell \leftarrow$  compute POD basis from new snapshots  $[y^{k+1}, p^{k+1}]$ 
15:  end if
16:   $\mu^{k+1}, y^{k+1}, \{\psi_j\}_{j=1}^\ell \leftarrow$  Algorithm 2: update control, state and POD basis if
    sufficient decrease condition is not fulfilled
17:   $p^{k+1} \leftarrow$  solve (adjoint equation) for  $\mu^{k+1}$  and  $y^{k+1}$  using ROM
18:   $\nabla J^{k+1} \leftarrow$  evaluate (reduced gradient) for  $\mu^{k+1}$ 
19:   $k \leftarrow k + 1$ 
20: end while
21:  $\mu \leftarrow \mu^k$ 
22:  $y \leftarrow y^k$ 

```

To obtain a reduced order model (ROM) for the state equations (2.1b)–(2.1d) and the adjoint equations (3.5) we solve them once using a finite element method (FEM) and then utilize a POD Galerkin scheme. When the parameter μ is updated an error indicator ρ is evaluated for the solution y (Algorithm 1, line 10). If this error indicator is too large (Algorithm 1, line 11) the algorithm initiates an update of the POD basis. The same strategy is applied in Armijo-backtracking (Algorithm 2). Note that for y and p a combined POD basis ψ is used. This strategy is proven to be effective since in the adjoint both variables y and p are present. The number of POD basis functions is denoted by ℓ . As error indicator the residual can be used. The residuals are computed by inserting the solution obtained by the ROM into the original problem discretized by the FEM. This estimates how good the solution is compared to a FEM discretization. Hence a decision can be made whether to trust

Algorithm 2 Modified line search with Armijo-backtracking strategy**Input:** $y^{k+1}, u^{k+1}, \{\psi_j\}_{j=1}^\ell, \mu^k, d^k, \ell, \varepsilon^{POD}, \tau, \sigma$ **Output:** $y^{k+1}, \mu^{k+1}, \{\psi_j\}_{j=1}^\ell$

```

1:  $\alpha \leftarrow 1$ 
2: while  $J(\mu^{k+1}) > J(\mu^k) - \alpha\sigma\langle\nabla J^k, d^k\rangle$  do
3:    $\alpha \leftarrow \alpha/2$ 
4:    $\mu^{k+1} \leftarrow \mu^k + \alpha d^k$ 
5:    $y^{k+1} \leftarrow$  solve (state equation) for  $\mu^{k+1}$  using ROM
6:    $\rho \leftarrow$  evaluate error indicator for  $y^{k+1}$ 
7:   if  $\rho > \varepsilon^{POD}$  then
8:      $y^{k+1} \leftarrow$  solve (state equation) for  $\mu^{k+1}$  using FEM
9:      $p^{k+1} \leftarrow$  solve (adjoint equation) for  $\mu^{k+1}$  and  $y^{k+1}$  using FEM
10:     $\{\psi_j\}_{j=1}^\ell \leftarrow$  compute POD basis from new snapshots  $[y^{k+1}, p^{k+1}]$ 
11:   end if
12: end while

```

the solution of the ROM or to recompute the solution for a given μ using FEM and update the POD basis. Since we are considering a nonlinear problem we utilize EIM [9, 22] to obtain an efficient reduced order model. In this case we can utilize a more efficient error estimator that does not require the assembling of the nonlinear system matrix. For details we refer the reader to [9, Proposition 4.1]. Note that when ever the state equation is solved also the basis for the EIM is updated.

The search direction d^k in line 7 is computed by using a Newton-CG method (e.g. [23], p. 169, or [16], p. 30) solving

$$(\nabla^2 J^k)d^k = -\nabla J^k,$$

where $\nabla^2 J^k$ is the Hessian representation of the cost functional J at iteration k . For the implementation we make use of the fact, that — as described in section 3.3, Remark 3.9 — we do not need to set up the Hessian matrix explicitly, but can compute the effect of the Hessian representation applied on a vector.

Algorithm 1 has the advantage that over the number of optimization iterations the POD basis is improved and taylored to the optimization problem. On the other hand, if the problem is not suited for a model order reduction method, the algorithm uses only FE solution. Hence convergence is guaranteed for the algorithm. As a remark, setting ε^{POD} to infinity disables the adaptive algorithm and we end up with an optimization algorithm that runs fully on the ROM.

6.2. The a-posteriori error estimator. In Algorithm 3 we give a short insight, how the a-posteriori error estimation is implemented. To evaluate the perturbation function ζ (cp. (4.5)), full system solves for the state equations and for the adjoint equations using FEM have to be done. For the computation of the smallest eigenvalue of the Hessian we will consider different numerical approaches, which will be introduced in the next section.

7. NUMERICAL EXPERIMENTS

Since the application of POD for order reduction of high dimensional models lacks a (strict) a-priori error estimation, at least an efficient a-posteriori error estimator would be appreciated. While the knowledge of rigorous a-priori error bounds

Algorithm 3 A-posteriori error estimator

Input: μ **Output:** ε_{ape}

- 1: $y \leftarrow$ solve (*state equation*) for μ using FEM
 - 2: $p \leftarrow$ solve (*adjoint equation*) for y using FEM
 - 3: $\zeta \leftarrow$ evaluate ζ concerning μ , y and p
 - 4: $\sigma_{\min} \leftarrow$ compute smallest eigenvalue of $\hat{J}''(\mu)$
 - 5: $\varepsilon_{\text{ape}} \leftarrow 2/\sigma_{\min} \|\zeta\|$
-

could prevent one from performing an expensive computation step, for a-posteriori error analysis this computation step has to be already done. Especially for that reason, this error estimation has to be efficient concerning computational time and memory and, of course, accuracy. In the following we denote by 'CPU time' the time measuring with Matlab function `cputime`, while 'SW time' denotes the so called "stop watch" time that is provided using Matlab functions `tic` and `toc`.

7.1. Settings. For numerical results we implemented the semilinear parabolic optimal control problem (2.1) for the 2D spatial domain $\Omega = (0, 1) \times (0, 1)$ and the time interval $[0, T]$ for $T = 1$. We influence the system dynamics partially by $m \in \mathbb{N}$ space and time independent control parameters $\mu_i \in \mathbb{R}$, $i = 1, \dots, m$, and define $\mu = (\mu_1, \dots, \mu_m) \in \mathbb{R}^m$. The control parameters μ_i are distributed to Ω by non-negative shape functions $b_1, \dots, b_m \in L^\infty(\Omega)$. We choose the weighting parameters $\kappa_i = |\Omega_i|$, $i = 1, \dots, m$. The admissible set for the control is determined by

$$\mathcal{D}_{\text{ad}} = \{\mu \in \mathbb{R}^m \mid \mu^a \leq \mu \leq \mu^b\}.$$

For solving the semilinear parabolic optimal control problem (2.1) numerically we applied a finite element method with standard piecewise linear ansatz functions in space and the implicit Euler scheme in time. The nonlinearity in the PDE is handled by a Newton method.

We compare the optimization concerning the usage of the High-Dimensional Model (HDM), the (standard) Reduced-Order Model (ROM) and the Adaptive Reduced Order Model (ADAPT). While in the ROM approach a POD-EIM basis is fixed once at the beginning, the adaptive approach performs a basis update when a certain residual-based error estimation criterion is fulfilled.

7.2. Run 1 (Validation of the error estimator). For a predication of the quality of the error estimation, we validate our approach by considering an example where the exact optimal solution is known. The optimal and (therefore also) desired states in the cost functional (2.1a) are given by

$$\begin{aligned} \bar{y}(t, \mathbf{x}) &= y_Q(t, \mathbf{x}) = \cos(4\pi x_1) \cos(4\pi t x_2), \\ y_\Omega(\mathbf{x}) &= y_Q(T, \mathbf{x}), \end{aligned}$$

with weightings $\alpha_Q = \alpha_\Omega = 1$. An insight into the characteristics of the desired state is shown for some time instances in Fig. 7.1. The desired control is set component-wise to

$$\mu_i^\circ = \sin\left(\frac{i-1}{m-1}5\pi\right)$$

for all $i = 1, \dots, m$, with regularization $\nu = 10^{-2}$. We set the control bounds $\mu^a = 0$ and $\mu^b = 1.5$, aware of the fact, that for this definition of \mathcal{D}_{ad} the computed control

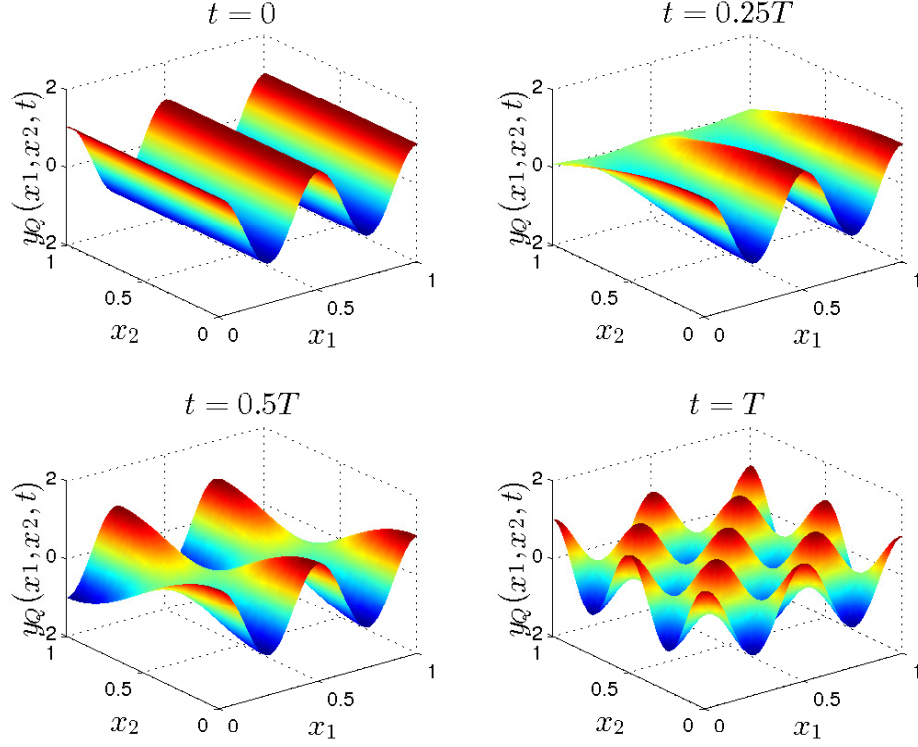


FIGURE 7.1. Run 1: Desired state solution $y_Q(t, \mathbf{x})$ on Ω for different time points t .

cannot reach the desired μ° exactly. The number of control parameters is set to $m = 25$.

The shape functions are given by

$$b_i(x_1, x_2, r_i, c_i, \omega_i) = \max(0, r_i^2 - ((x_1 - c_{i,1})^2 + (x_2 - c_{i,2})^2)) \frac{\omega_i}{r_i^2},$$

with radius $r_i \in \mathbb{R}^+$, center point $c_i \in \mathbb{R}^2$ and a maximum value scaling $\omega_i \in \mathbb{R}$ for all $i = 1, \dots, m$, see Fig. 7.2. The right-hand side f and initial function y_o in the semilinear parabolic differential equation (2.1b)-(2.1d) are set to

$$\begin{aligned} f(t, \mathbf{x}) = & -4 \cos(4\pi x_1) \sin(4\pi t x_2) x_2 \\ & - 16 \cos(4\pi x_1) \cos(4\pi t x_2) (1 + t^2) \\ & + \sinh \left(y_Q(t, \mathbf{x}) \sum_{i=1}^3 \mu_i^\circ b_i(x_1, x_2) \right), \end{aligned}$$

and

$$y_o(\mathbf{x}) = \cos(4\pi x_1).$$

We discretize the spatial domain Ω by a total number of $N_{\mathbf{x}} = 2601$ discretization points and consider the time domain $[0, T]$ for $T = 1$ at $N_t = 200$ discrete time points.

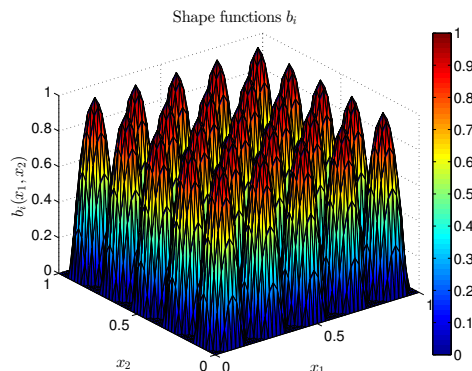


FIGURE 7.2. Run 1: Equally distributed parameter control influence for $m = 25$ shape functions.

	HDM	ROM	ADAPT
J	1.991e-02	2.163e-02	2.163e-02
#Iter	5	5	5
∅ Time per iter [s]	38.0	7.2	7.8
SW Time [s]	206.0	60.3	62.4
CPU Time [s]	530.0	149.1	167.7

TABLE 7.1. Run 1: Performance of the three methods.

We set as initial guess for the control parameter a random value

$$\mu_i^{(0)} \in [0, 1],$$

for each component $i = 1, \dots, m$. This parameter control is also utilized for the (initial) POD-EIM basis computation parameter $\mu_{bc} = \mu^{(0)}$ in case of ROM and ADAPT.

An overview of the performance of the three methods can be found in Tab. 7.1. Both methods ROM and ADAPT lead here to a total speed-up factor concerning optimization time of ≥ 3 , even when the average iteration time shows a speed-up of factor 5. This is due to the fact, that initially in case of ROM and ADAPT an computational expensive FE-Snapshot solve for the state and the adjoint equation has to be done for the POD (and EIM) basis computation, which has to be counted additionally. Despite the fact that we can use this computation as a first iterate, it compensates partially the advantage concerning optimization time, especially when only few iterations have to be done for optimizing. As can be seen in Tab. 7.2 and in Figure 7.3, and Figure 7.4 all three methods lead to good approximations of μ° . Please note here, that μ° was not reachable exactly, due to the control bounds.

For the (additional) a-posteriori error estimation in case of ROM, we compare the obtained error estimates for four different ways of smallest eigenvalue computation:

- (1) 'minLM': Determining the minimum of all eigenvalues computed by Matlab function `eigs` applied on the Hessian-vector computation with option 'Largest Magnitude'.

NEWT-CG	
$\ \mu^\circ - \mu^h\ _{\mathbb{R}^m}$	6.029e-01
$\ \mu^\circ - \mu^\ell\ _{\mathbb{R}^m}$	6.029e-01
$\ \mu^\circ - \mu^A\ _{\mathbb{R}^m}$	6.029e-01
$\ \mu^h - \mu^\ell\ _{\mathbb{R}^m}$	1.619e-04
$\ \mu^h - \mu^A\ _{\mathbb{R}^m}$	1.619e-04
$\ \mu^\ell - \mu^A\ _{\mathbb{R}^m}$	6.939e-18

TABLE 7.2. Run 1: Quality of HDM (μ^h), ROM (μ^ℓ) and ADAPT (μ^A) approximations, compared to μ° .

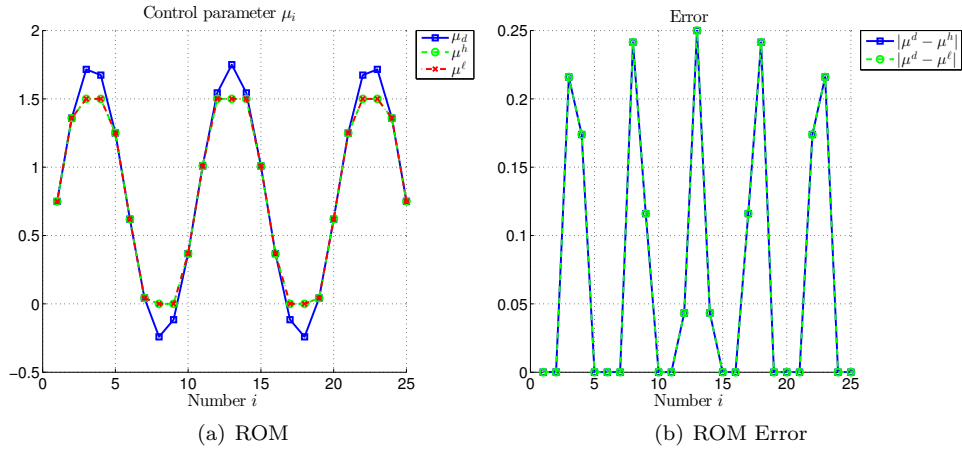


FIGURE 7.3. Run 1: Comparing the solutions for μ (ROM).

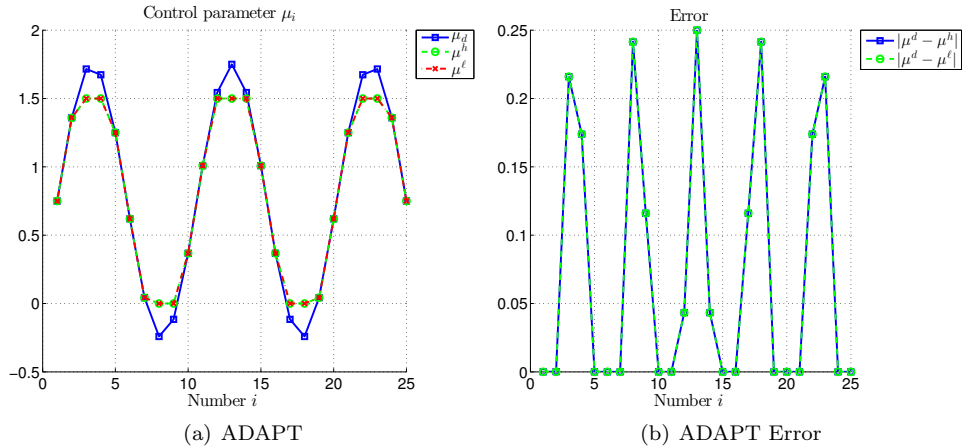


FIGURE 7.4. Run 1: Comparing the solutions for μ (ADAPT).

	SW Time [s]	CPU Time [s]
Solver FE State	18.5	48.4
Solver FE Adjoint	4.3	11.9
Perturbation ζ	0.5	0.7

TABLE 7.3. Run 1: Computation times.

	minLM	SM	Hess	Sens
SW Time [s]	220.7	163.7	88.7	56.5
CPU Time [s]	555.1	417.2	234.5	141.0
σ_{\min}	9.9998e-02	1.0000e-01	9.9998e-02	9.9998e-02
$\ \zeta\ _{\mathbb{R}^m}$	1.6282e-05	1.6282e-05	1.6282e-05	1.6282e-05
$\varepsilon_{\text{ape}} = \frac{2}{\sigma_{\min}} \ \zeta\ _{\mathbb{R}^m}$	3.2565e-04	3.2564e-04	3.2565e-04	3.2565e-04

TABLE 7.4. Run 1: A-posteriori error estimation.

- (2) 'SM': Computing only the minimum eigenvalue by Matlab function `eigs` applied on the Hessian-vector computation wrapped in a linear equation solver with option '`Smallest Magnitude`'.
- (3) 'Hess': Setting up the full Hessian matrix by the Hessian-vector computation applied on a basis of \mathcal{D}_{ad} and using Matlab `eig`.
- (4) 'Sens': Setting up the full Hessian matrix via the sensitivity approach by computing the directional derivatives for a basis of \mathcal{D}_{ad} (cf. [11]) and using Matlab `eig`.

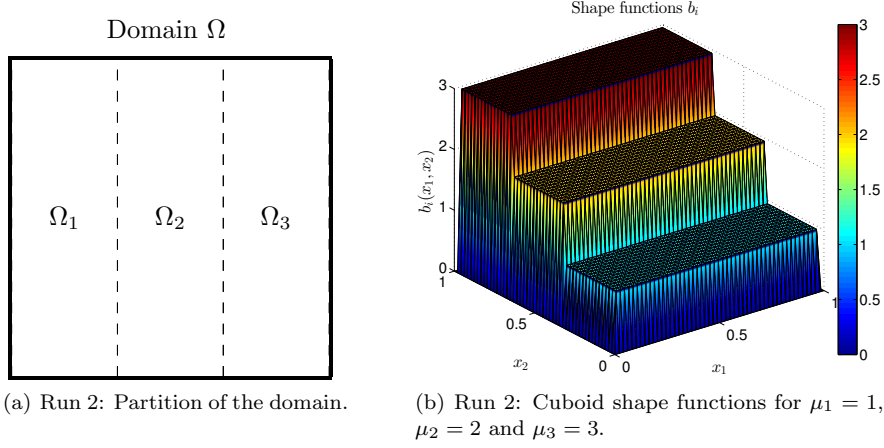
The times for solving state and adjoint with FE and the evaluation of perturbation function ζ are given in Tab. 7.3. The times and results for computing the smallest eigenvalue and the a-posteriori error estimator are presented in Tab. 7.4.

7.3. Run 2 (simplified battery model). This numerical example is inspired by a simplified version of the PDE systems arising in battery equations, which was handled in [8, 28] as a model for the lithium-ion concentrations. Let us also refer to [14], where a parameter identification problem is considered for a different battery model for the lithium-ion concentrations. The spatial discretization of Ω is set to $N_{\mathbf{x}} = 10201$ while the time domain $[0, T]$ for $T = 1$ is discretized in $N_t = 2500$ time points. The number of control parameters is set to $m = 3$. We consider the domain to be partitioned in 3 equal sized subdomains $\Omega = \Omega_1 \cup \Omega_2 \cup \Omega_3$, where the control parameters μ_i , $i = 1, \dots, m$, are applied by ‘‘cuboid’’ shape functions

$$b_i(x_1, x_2) = \begin{cases} 1 & \text{if } (x_1, x_2) \in \Omega_i, \\ 0 & \text{else.} \end{cases}$$

as shown in Fig. 7.3. In the cost functional (2.1a) we set the desired control parameter vector

$$\mu^\circ = (\mu_1^\circ, \dots, \mu_m^\circ) = 0$$



(a) Run 2: Partition of the domain.

(b) Run 2: Cuboid shape functions for $\mu_1 = 1$, $\mu_2 = 2$ and $\mu_3 = 3$.

with regularization $\nu = 1.0e - 2$ and weighting $\kappa_i = 1$, for $i = 1, \dots, m$, while the desired states are given by

$$y_Q(t, \mathbf{x}) = \cos(2\pi x_1) \cos(2\pi t x_2)$$

and

$$y_\Omega(\mathbf{x}) = \cos(2\pi x_1) \cos(2\pi T x_2),$$

with weightings $\alpha_Q = \alpha_\Omega = 100$. An insight into the characteristics of the desired state solution in Q is shown for some time instances in Fig. 7.5.

The right-hand side f and initial function y_o in the semilinear parabolic differential equation (2.1b)-(2.1d) are set to

$$\begin{aligned} f(t, \mathbf{x}) = & -2 \cos(2\pi x_1) \sin(2\pi t x_2) x_2 \\ & - 4 \cos(2\pi x_1) \cos(2\pi t x_2) (1 + t^2) \\ & + \sinh \left(y_Q(t, \mathbf{x}) \sum_{i=1}^3 \mu_i^f b_i(x_1, x_2) \right) \end{aligned}$$

with $\mu^f = (\mu_1^f, \mu_2^f, \mu_3^f) = (4.0, 4.5, 5.0)$, and

$$y_o(\mathbf{x}) = \cos(2\pi x_1).$$

For the admissible parameter set \mathcal{D}_{ad} we fix the parameter control bounds at $\mu^a = 0$ and $\mu^b = 7.0$.

We compare the optimization concerning the usage of the High-Dimensional FE-Model (HDM), the (standard) Reduced-Order Model (ROM) and the Adaptive Reduced Order Model (ADAPT). In the case of the standard ROM additionally the a-posteriori error estimator (APE) is taken into account. We apply for the smallest eigenvalue computation only the sensitivity approach ('Sens') as introduced before, since for this setting it turned out to be the fastest without observable loss in accuracy. While in the ROM approach the POD as well as the EIM basis is fixed once at the beginning, the adaptive approach performs basis updates when a certain residual-based error estimation criterion is fulfilled.

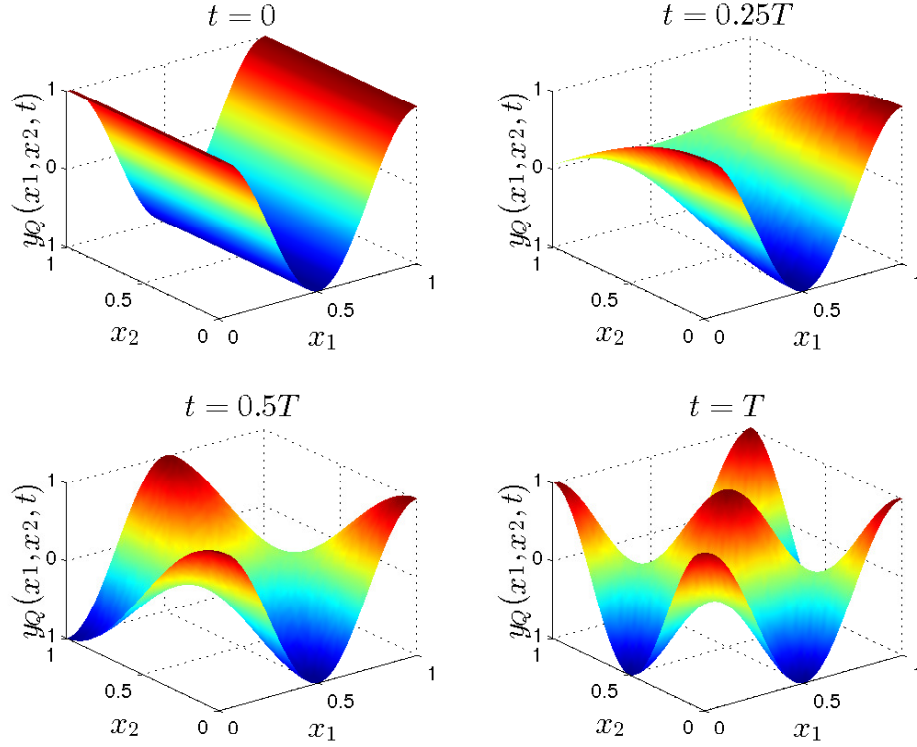


FIGURE 7.5. Run 2: Desired state solution $y_Q(t, \mathbf{x})$ on Ω for different time points t .

We set as initial guess for the control parameters

$$\mu_i^{(0)} = 0, \quad i = 1, \dots, m.$$

The snapshots for generating the (initial) POD and EIM basis in case of ROM and ADAPT are computed for control parameters

$$\mu_i^{\text{bc}} = 3.0, \quad i = 1, \dots, m.$$

The POD basis range was set to a number of 15 basis functions while for the EIM approximation of the nonlinearity up to a tolerance of $1.0e-12$ we get a number of 144 basis functions. The numerical optimization was done by utilizing a projected Newton-CG approach; see [16], for instance.

For ensuring numerical convergence and stability in the ROM case, one has to deal with the dynamics of the hyperbolic sine (and cosine), see Fig. 7.6, arising in the state solver, where the nonlinear equation is solved by a Newton method. This is handled by introducing bounds for the argument of the hyperbolic sine.

As we can see in Tab. 7.5 for the performance concerning optimization time we gain a factor around 4 for ROM and a factor around 3 for ADAPT, compared to the optimization utilizing the high-dimensional model HDM.

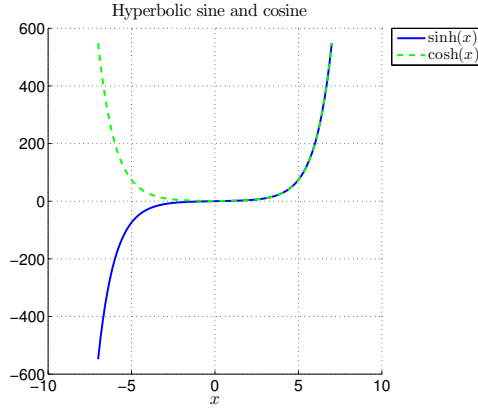


FIGURE 7.6. Run 2: Dynamics of the nonlinear functions sinh and cosh.

	HDM	ROM	ADAPT
J	4.548e-01	4.604e-01	4.556e-01
#Iter	12	13	13
OPT SW Time [s]	8570	2146	2843
OPT CPU Time [s]	8562	2144	2840
APE SW Time [s]	—	651	—
APE CPU Time [s]	—	650	—

TABLE 7.5. Run 2: Performance of the three methods.

$\ \mu^f - \mu^h\ _{\mathbb{R}^m}$	1.797e-01
$\ \mu^f - \mu^\ell\ _{\mathbb{R}^m}$	1.829e-01
$\ \mu^f - \mu^A\ _{\mathbb{R}^m}$	1.802e-01
$\ \mu^h - \mu^\ell\ _{\mathbb{R}^m}$	1.258e-02
$\ \mu^h - \mu^A\ _{\mathbb{R}^m}$	3.597e-03
$\ \mu^\ell - \mu^A\ _{\mathbb{R}^m}$	1.286e-02

TABLE 7.6. Run 2: Quality of HDM (μ^h), ROM (μ^ℓ) and ADAPT (μ^A) approximations, compared to μ^f .

All three methods lead to a good approximation of the control parameter μ^f influencing the right-hand side of the semilinear parabolic PDE, see Tab. 7.6. Furthermore we can state, that the ROM as well as the ADAPT approach leads to good approximations of the optimal control parameters computed with HDM.

As can be seen in Tab. 7.7 the a-posteriori error estimator ε_{ape} gives a quite good and sharp upper bound concerning the norm of the distance $\mu^h - \mu^\ell$ that we effectively have to consider, when the HDM optimization is replaced by a faster ROM optimization. Despite the fact that we have to invest additional time after the optimization process for performing the a-posteriori error estimation, we finally end

Norm of perturbation $\ \zeta\ _{\mathbb{R}^m}$	3.678e-03
Smallest Eigenvalue σ_{\min}	2.516e-01
$\varepsilon_{\text{ape}} := \frac{2}{\sigma_{\min}} \ \zeta\ _{\mathbb{R}^m}$	2.924e-02

TABLE 7.7. Run 2: A-posteriori error estimator.

	SW Time [s]	CPU Time [s]
Solver FE State	370.1	369.8
Solver FE Adjoint	80.7	80.6
Perturbation ζ	2.8	2.8
Smallest Eigenvalue σ_{\min}	197.3	197.2

TABLE 7.8. Run 2: Computational times for the evaluation of the a-posteriori error estimator.

up with a speed-up factor of 3 compared to the HDM optimization. Furthermore, in this case, we catch up with our ADAPT approach. The single computation times for the a-posteriori error estimation can be found in Tab. 7.8.

APPENDIX: PROOFS

Proof of Proposition 2.4. Suppose that $y^1 = \mathcal{G}(\mu^1)$ and $y^2 = \mathcal{G}(\mu^2)$. Then $y = y^1 - y^2$ is the weak solution to

$$\begin{aligned}
 (1) \quad & y_t - \Delta y + \sinh\left(y^1 \sum_{i=1}^m \mu_i^1 b_i\right) - \sinh\left(y^2 \sum_{i=1}^m \mu_i^2 b_i\right) = 0 \quad \text{in } Q, \\
 & \frac{\partial y}{\partial n} = 0 \quad \text{on } \Sigma, \\
 & y(0, \cdot) = 0 \quad \text{in } \Omega.
 \end{aligned}$$

Writing

$$\begin{aligned}
 & \sinh\left(y^1 \sum_{i=1}^m \mu_i^1 b_i\right) - \sinh\left(y^2 \sum_{i=1}^m \mu_i^2 b_i\right) \\
 &= \sinh\left(y^1 \sum_{i=1}^m \mu_i^1 b_i\right) - \sinh\left(y^2 \sum_{i=1}^m \mu_i^1 b_i\right) + \sinh\left(y^2 \sum_{i=1}^m \mu_i^1 b_i\right) - \sinh\left(y^2 \sum_{i=1}^m \mu_i^2 b_i\right)
 \end{aligned}$$

we apply the mean value theorem twice times

$$\begin{aligned}
 & \sinh\left(y^1 \sum_{i=1}^m \mu_i^1 b_i\right) - \sinh\left(y^2 \sum_{i=1}^m \mu_i^1 b_i\right) \\
 &= \int_0^1 \sum_{i=1}^m \mu_i^1 b_i \cosh\left(\left(y^2 + sy\right) \sum_{i=1}^m \mu_i^1 b_i\right) ds y, \\
 & \sinh\left(y^2 \sum_{i=1}^m \mu_i^1 b_i\right) - \sinh\left(y^2 \sum_{i=1}^m \mu_i^2 b_i\right) \\
 &= \int_0^1 y^2 \cosh\left(y^2 \sum_{i=1}^m (\mu_i^2 + s\mu_i) b_i\right) ds \sum_{i=1}^m \mu_i b_i.
 \end{aligned}$$

We define the functions

$$\begin{aligned}\alpha(t, \mathbf{x}) &= \sum_{i=1}^m \mu_i^1 b_i(\mathbf{x}) \int_0^1 \cosh\left(\left(y^2(t, \mathbf{x}) + sy(t, \mathbf{x})\right) \sum_{i=1}^m \mu_i^1 b_i(\mathbf{x})\right) ds, \\ \beta(t, \mathbf{x}) &= -y^2(t, \mathbf{x}) \int_0^1 \cosh\left(y^2(t, \mathbf{x}) \sum_{i=1}^m (\mu_i^2 + s\mu_i) b_i(\mathbf{x})\right) ds.\end{aligned}$$

Then, $\alpha, \beta \in L^\infty(Q)$, $\alpha \geq 0$ in Q a.e. and α is monotonically increasing. In particular,

$$\begin{aligned}\|\beta\|_{L^r(Q)}^r &= \int_0^T \int_\Omega |\beta(t, \mathbf{x})|^r d\mathbf{x} dt \\ &= \int_0^T \int_\Omega \left| \int_0^1 y^2(t, \mathbf{x}) \cosh\left(y^2(t, \mathbf{x}) \sum_{i=1}^m (\mu_i^2 + s\mu_i) b_i(\mathbf{x})\right) ds \right|^r d\mathbf{x} dt \\ &\leq \|y^2\|_{L^\infty(Q)} \int_0^T \int_\Omega \left| \cosh\left(m \max_{1 \leq i \leq m} (\mu_i^b, \|b_i\|_{L^\infty(\Omega)}) \|y^2\|_{L^\infty(Q)}\right) \right|^r d\mathbf{x} dt.\end{aligned}$$

From (2.4) we infer that there is a constant $C_1 > 0$ independent of y^1, y^2 and μ^1, μ^2 so that $\|\beta\|_{L^r(Q)} \leq C_1$. Summarizing, we derive from (.1)

$$\begin{aligned}y_t - \Delta y + \alpha(t, \mathbf{x})y &= \beta(t, \mathbf{x}) \sum_{i=1}^m \mu_i b_i && \text{in } Q, \\ \frac{\partial y}{\partial n} &= 0 && \text{on } \Sigma, \\ y(0, \cdot) &= 0 && \text{in } \Omega.\end{aligned}\tag{.2}$$

Due to [26, Theorem 5.5] there exists a unique solution to (.2) and there is a constant $C_2 > 0$ independent of α so that

$$\begin{aligned}\|y^1 - y^2\|_Y &= \|y\|_Y \leq C_2 \left\| \beta \sum_{i=1}^m \mu_i b_i \right\|_{L^r(Q)} \leq C_2 \left\| \sum_{i=1}^m \mu_i b_i \right\|_{L^\infty(\Omega)} \|\beta\|_{L^r(Q)} \\ &\leq C_1 C_2 \sum_{i=1}^m \|\mu_i\| \|b_i\|_{L^\infty(\Omega)} \leq C_3 |\mu|_2 = C_3 |\mu^1 - \mu^2|_2\end{aligned}$$

with $C_3 = C_1 C_2 (\sum_{i=1}^m \|b_i\|_{L^\infty(\Omega)}^2)^{1/2} > 0$ independent of y^1, y^2 and μ^1, μ^2 . \square

Proof of Proposition 3.1. Let $\bar{y} = \mathcal{G}(\bar{\mu})$ and $\tilde{y} = \mathcal{G}(\bar{\mu} + \mu)$. Then the difference $\tilde{y} - \bar{y}$ satisfies the parabolic problem

$$\begin{aligned}(\tilde{y} - \bar{y})_t - \Delta(\tilde{y} - \bar{y}) + \sinh\left(\tilde{y} \sum_{i=1}^m (\bar{\mu}_i + \mu_i) b_i\right) - \sinh\left(\bar{y} \sum_{i=1}^m \bar{\mu}_i b_i\right) &= 0 && \text{in } Q, \\ \frac{\partial(\tilde{y} - \bar{y})}{\partial n} &= 0 && \text{on } \Sigma, \\ (\tilde{y} - \bar{y})(0, \cdot) &= 0 && \text{in } \Omega.\end{aligned}\tag{.3}$$

Notice that

$$\begin{aligned}\sinh\left(\tilde{y} \sum_{i=1}^m (\bar{\mu}_i + \mu_i) b_i\right) - \sinh\left(\bar{y} \sum_{i=1}^m \bar{\mu}_i b_i\right) &= \sinh\left(\tilde{y} \sum_{i=1}^m (\bar{\mu}_i + \mu_i) b_i\right) \\ &- \sinh\left(\tilde{y} \sum_{i=1}^m \bar{\mu}_i b_i\right) + \sinh\left(\tilde{y} \sum_{i=1}^m \bar{\mu}_i b_i\right) - \sinh\left(\bar{y} \sum_{i=1}^m \bar{\mu}_i b_i\right).\end{aligned}\tag{.4}$$

Since $\tilde{y} \in C(\overline{Q})$ and $b_i \in C(\overline{\Omega})$ hold, we can apply Taylor's theorem to get

$$(5) \quad \sinh\left(\tilde{y} \sum_{i=1}^m (\bar{\mu}_i + \mu_i) b_i\right) - \sinh\left(\tilde{y} \sum_{i=1}^m \bar{\mu}_i b_i\right) = \tilde{y} \cosh\left(\tilde{y} \sum_{i=1}^m \bar{\mu}_i b_i\right) \sum_{i=1}^m \mu_i b_i + r_\mu$$

with a residual $r_\mu \in L^\infty(Q)$ satisfying $\|r_\mu\|_{L^\infty(Q)}/|\mu|_2 \rightarrow 0$ for $|\mu|_2 \rightarrow 0$. Note that

$$\tilde{y} \cosh\left(\tilde{y} \sum_{i=1}^m \bar{\mu}_i b_i\right) = \bar{y} \cosh\left(\bar{y} \sum_{i=1}^m \bar{\mu}_i b_i\right) + \tilde{y} \cosh\left(\tilde{y} \sum_{i=1}^m \bar{\mu}_i b_i\right) - \bar{y} \cosh\left(\bar{y} \sum_{i=1}^m \bar{\mu}_i b_i\right).$$

Hence, we obtain from (.3), (.4) and (.5)

$$(6) \quad \begin{aligned} & (\tilde{y} - \bar{y})_t - \Delta(\tilde{y} - \bar{y}) + \sinh\left(\tilde{y} \sum_{i=1}^m (\bar{\mu}_i + \mu_i) b_i\right) - \sinh\left(\bar{y} \sum_{i=1}^m \bar{\mu}_i b_i\right) \\ &= (\tilde{y} - \bar{y})_t - \Delta(\tilde{y} - \bar{y}) + \bar{y} \cosh\left(\bar{y} \sum_{i=1}^m \bar{\mu}_i b_i\right) \sum_{i=1}^m \mu_i b_i + r_\mu \\ &+ \left(\tilde{y} \cosh\left(\tilde{y} \sum_{i=1}^m \bar{\mu}_i b_i\right) - \bar{y} \cosh\left(\bar{y} \sum_{i=1}^m \bar{\mu}_i b_i\right)\right) \sum_{i=1}^m \mu_i b_i \\ &+ \sinh\left(\tilde{y} \sum_{i=1}^m \bar{\mu}_i b_i\right) - \sinh\left(\bar{y} \sum_{i=1}^m \bar{\mu}_i b_i\right) \quad \text{in } Q. \end{aligned}$$

Recall that the Nemytskii operator $\Phi_{\bar{\mu}}$ is continuously Fréchet-differentiable. Thus, by Taylor expansion

$$(7) \quad \sinh\left(\tilde{y} \sum_{i=1}^m \bar{\mu}_i b_i\right) - \sinh\left(\bar{y} \sum_{i=1}^m \bar{\mu}_i b_i\right) = \sum_{i=1}^m \bar{\mu}_i b_i \cosh\left(\bar{y} \sum_{i=1}^m \bar{\mu}_i b_i\right) (\tilde{y} - \bar{y}) + r_y,$$

where the residual $r_y \in L^\infty(Q)$ satisfies $\|r_y\|_{L^\infty(Q)}/\|y\|_{L^\infty(Q)} \rightarrow 0$ for $\|y\|_{L^\infty(Q)} \rightarrow 0$. Summarizing, we infer from (.6) and (.7) that

$$\begin{aligned} & (\tilde{y} - \bar{y})_t - \Delta(\tilde{y} - \bar{y}) + \sinh\left(\tilde{y} \sum_{i=1}^m (\bar{\mu}_i + \mu_i) b_i\right) - \sinh\left(\bar{y} \sum_{i=1}^m \bar{\mu}_i b_i\right) \\ &= (\tilde{y} - \bar{y})_t - \Delta(\tilde{y} - \bar{y}) + \bar{y} \cosh\left(\bar{y} \sum_{i=1}^m \bar{\mu}_i b_i\right) \sum_{i=1}^m \mu_i b_i + r_\mu \\ &+ \left(\tilde{y} \cosh\left(\tilde{y} \sum_{i=1}^m \bar{\mu}_i b_i\right) - \bar{y} \cosh\left(\bar{y} \sum_{i=1}^m \bar{\mu}_i b_i\right)\right) \sum_{i=1}^m \mu_i b_i \\ &+ \sum_{i=1}^m \bar{\mu}_i b_i \cosh\left(\bar{y} \sum_{i=1}^m \bar{\mu}_i b_i\right) (\tilde{y} - \bar{y}) + r_y \quad \text{in } Q. \end{aligned}$$

Let $\tilde{y} - \bar{y} = y + \hat{y} + y_r$, where y solves (3.1), \hat{y} is the solution to

$$(8) \quad \begin{aligned} \hat{y}_t - \Delta \hat{y} &= \left(\bar{y} \cosh\left(\bar{y} \sum_{i=1}^m \bar{\mu}_i b_i\right) - \tilde{y} \cosh\left(\tilde{y} \sum_{i=1}^m \bar{\mu}_i b_i\right)\right) \sum_{i=1}^m \mu_i b_i && \text{in } Q, \\ \frac{\partial \hat{y}}{\partial n} &= 0 && \text{on } \Sigma, \\ \hat{y}(0, \cdot) &= 0 && \text{in } \Omega \end{aligned}$$

and y_r is the solution to

$$(9) \quad \begin{aligned} (\tilde{y}_r)_t - \Delta \tilde{y}_r &= -r_y - r_\mu && \text{in } Q, \\ \frac{\partial \tilde{y}_r}{\partial n} &= 0 && \text{on } \Sigma, \\ y_r(0, \cdot) &= 0 && \text{in } \Omega. \end{aligned}$$

Notice that the Nemytskii operator

$$\Phi : L^\infty(Q) \rightarrow L^\infty(Q), \quad (\Phi v)(t, \mathbf{x}) = v(t, \mathbf{x}) \cosh \left(v(t, \mathbf{x}) \sum_{i=1}^m \bar{\mu}_i b_i(\mathbf{x}) \right)$$

is Fréchet-differentiable. From (2.4) it follows that there exists a $C_1 > 0$ independent of \tilde{y} , \bar{y} , μ and $\bar{\mu}$ such that $\|\tilde{y}\|_{L^\infty(Q)} \leq C_1$ and $\|\bar{y}\|_{L^\infty(Q)} \leq C_1$. Since Φ is locally Lipschitz-continuous, there is a constant $C_2 > 0$ depending on C_1

$$\begin{aligned} & \left\| \left(\bar{y} \cosh \left(\bar{y} \sum_{i=1}^m \bar{\mu}_i b_i \right) - \tilde{y} \cosh \left(\tilde{y} \sum_{i=1}^m \bar{\mu}_i b_i \right) \right) \sum_{i=1}^m \mu_i b_i \right\|_{L^\infty(Q)} \\ & \leq \left\| \left(\bar{y} \cosh \left(\bar{y} \sum_{i=1}^m \bar{\mu}_i b_i \right) - \tilde{y} \cosh \left(\tilde{y} \sum_{i=1}^m \bar{\mu}_i b_i \right) \right) \right\|_{L^\infty(Q)} \sum_{i=1}^m |\mu_i| \|b_i\|_{L^\infty(\Omega)} \\ & \leq C_2 \|\bar{y} - \tilde{y}\|_{L^\infty(Q)} \left(\sum_{i=1}^m \|b_i\|_{L^\infty(\Omega)}^2 \right)^{1/2} |\mu|_2 = C_3 \|\mathcal{G}(\bar{\mu} + \mu) - \mathcal{G}(\bar{\mu})\|_{L^\infty(Q)} |\mu|_2 \\ & = \mathcal{O}(|\mu|_2^2), \end{aligned}$$

where we have used Proposition 2.4 and $C_3 = C_2(\sum_{i=1}^m \|b_i\|_{L^\infty(\Omega)}^2)^{1/2}$. Now we apply Theorem 2.2 to (.8) and use (2.4). Then, it follows that

$$(10) \quad \|\hat{y}\|_{\mathcal{Y}} = o(|\mu|_2).$$

Again, by Theorem 2.2 there is a constant $C_5 > 0$ such that

$$(11) \quad \|y_r\|_{\mathcal{Y}} \leq C_5 \|r_y + r_\mu\|_{L^r(Q)} = o(|\mu|_2),$$

where we have used (2.4). From (.10) and (.11) we conclude

$$\|\mathcal{G}(\bar{\mu} + \mu) - \mathcal{G}(\mu) - \mathcal{G}'(\bar{\mu})\mu\|_{L^\infty(Q)} = \|\tilde{y} - \bar{y} - y\|_{L^\infty(Q)} = \|\hat{y} - y_r\|_{L^\infty(Q)} = o(|\mu|_2).$$

It follows from [26, p. 213] that

$$\|y\|_{\mathcal{Y}} \leq C \left\| \bar{\beta} \sum_{i=1}^m \mu_i b_i \right\|_{L^r(Q)}.$$

Let $C_6 > 0$ be chosen such that $\|\bar{\beta}\|_{L^\infty(Q)} \max_{1 \leq i \leq m} \|b_i\|_{L^\infty(\Omega)} \leq C_6$ hold. Then,

$$\begin{aligned} \left\| \bar{\beta} \sum_{i=1}^m \mu_i b_i \right\|_{L^r(Q)} &= \left(\int_0^T \int_\Omega \left| \bar{\beta}(t, \mathbf{x}) \sum_{i=1}^m \mu_i b_i(\mathbf{x}) \right|^r d\mathbf{x} dt \right)^{1/r} \\ &\leq \|\bar{\beta}\|_{L^\infty(Q)} \max_{1 \leq i \leq m} \|b_i\|_{L^\infty(\Omega)} (T|\Omega|)^{1/r} \sum_{i=1}^m |\mu_i| \\ &= C |\mu|_2 \end{aligned}$$

where $C = \sqrt{m} C_6 (T|\Omega|)^{1/r}$. Notice that C depends on T , $|\Omega|$, $\bar{\beta}$, m and the b_i 's. \square

Proof of Proposition 3.6. By Assumption 1-1) the Nemytskii operator Φ_μ is twice continuously Fréchet-differentiable from $L^\infty(Q)$ to $L^\infty(Q)$ for any $\mu \in \mathcal{D}_{ad}$ (see Section 2.2). Then, the proof follows from Theorems 5.15 and 5.16 in [26, pp. 227-229]. \square

Proof of Proposition 5.2. Let μ belong to \mathcal{D}_{ad} . Inserting the Galerkin ansatz for $y^\ell(t) = \sum_{j=1}^{\ell} y_j^\ell(t) \psi_j$ into the weak formulation (5.3) we get the nonlinear system of ordinary differential equations

$$(12) \quad \begin{aligned} \mathbf{M} \dot{\mathbf{y}}^\ell(t) + \mathbf{S} \mathbf{y}^\ell(t) + \mathbf{d}(\mathbf{y}(t), \mu) &= \mathbf{f}(t) \quad \text{f.a.a. } t \in (0, T], \\ \mathbf{M} \mathbf{y}^\ell(0) &= \mathbf{y}_\circ, \end{aligned}$$

where we have set

$$\begin{aligned} \mathbf{M} &= \left(\left(\int_{\Omega} \psi_j \psi_i \, d\mathbf{x} \right) \right)_{1 \leq i, j \leq \ell}, & \mathbf{f}(t) &= \left(\int_{\Omega} f(t) \psi_i \, d\mathbf{x} \right)_{1 \leq i \leq \ell}, \\ \mathbf{S} &= \left(\left(\int_{\Omega} \nabla \psi_j \cdot \nabla \psi_i \, d\mathbf{x} \right) \right)_{1 \leq i, j \leq \ell}, & \mathbf{y}^\ell(t) &= (y_i^\ell(t))_{1 \leq i \leq \ell}, \\ \mathbf{d}(\mathbf{y}(t), \mu) &= \left(\int_{\Omega} d(\cdot, \mathbf{y}^\ell(t), \mu) \psi_i \, d\mathbf{x} \right)_{1 \leq i \leq \ell}, & \mathbf{y}_\circ &= \left(\int_{\Omega} y_\circ \psi_i \, d\mathbf{x} \right)_{1 \leq i \leq \ell}. \end{aligned}$$

For any ℓ the matrices \mathbf{M} and \mathbf{S} are symmetric and positive definite. Let us introduce the locally Lipschitz-continuous nonlinear mapping $\mathcal{F}(t, \mathbf{y}^\ell(t), \mu) = \mathbf{M}^{-1}(\mathbf{f}(t) - \mathbf{S} \mathbf{y}^\ell(t) - \mathbf{d}(\mathbf{y}(t), \mu))$. Then, (12) can be expressed as

$$(13) \quad \dot{\mathbf{y}}^\ell(t) = \mathcal{F}(\mathbf{y}^\ell(t), \mu) \quad \text{f.a.a. } t \in (0, T], \quad \mathbf{y}^\ell(0) = \mathbf{M}^{-1} \mathbf{y}_\circ.$$

It follows from the Picard-Lindelöf theorem that there exists a T_* so that (13) has a unique solution $\mathbf{y}^\ell \in H^1(0, T_*; \mathbb{R}^\ell)$. Next we prove that we can choose $T_* = T$. Multiplying (5.3a) by $y_j^\ell(t)$ and summation from $j = 1$ to ℓ yields

$$\frac{1}{2} \frac{d}{dt} \|\mathbf{y}^\ell(t)\|_H^2 + \int_{\Omega} |\nabla \mathbf{y}^\ell(t)|_2^2 \, d\mathbf{x} + \int_{\Omega} d(\cdot, \mathbf{y}^\ell(t), \mu) \mathbf{y}^\ell(t) \, d\mathbf{x} \leq \|f(t)\|_H \|\mathbf{y}^\ell(t)\|_H$$

f.a.a. $t \in [0, T]$. From $\sum_{l=1}^m \mu_l b_l \geq 0$ in Ω a.e. we deduce that

$$d(\mathbf{x}, \tilde{\mathbf{y}}, \mu) \tilde{\mathbf{y}} = \sinh \left(\tilde{\mathbf{y}} \sum_{l=1}^m \mu_l b_l \right) \tilde{\mathbf{y}} \geq 0 \quad \text{f.a.a. } \mathbf{x} \in \Omega \text{ and for all } \tilde{\mathbf{y}} \in \mathbb{R}.$$

Thus,

$$(14) \quad \frac{1}{2} \frac{d}{dt} \|\mathbf{y}^\ell(t)\|_H^2 + \int_{\Omega} |\nabla \mathbf{y}^\ell(t)|_2^2 \, d\mathbf{x} \leq \|f(t)\|_H \|\mathbf{y}^\ell(t)\|_H.$$

Utilizing the Gronwall lemma we infer that

$$\|\mathbf{y}^\ell\|_{H^1(0, T; V^\ell)} \leq C (\|f\|_{L^2(0, T; H)} + \|y_\circ\|_{L^2(\Omega)})$$

for a constant $C > 0$, which implies existence of a solution \mathbf{y}^ℓ to (12) on $[0, T]$. \square

Proof of Proposition 5.3. Let us introduce the Ritz projection $\mathcal{P}^\ell : V \rightarrow V^\ell = \text{span} \{\psi_1, \dots, \psi_\ell\}$ as follows:

$$\langle \mathcal{P}^\ell \varphi, \psi \rangle_V = \langle \varphi, \psi \rangle_V \quad \text{for all } \psi \in V^\ell.$$

It follows from the Riesz theorem that the linear operator \mathcal{P} is well-defined and bounded. We decompose the error as

$$y(t) - y^\ell(t) = y(t) - \mathcal{P}^\ell y(t) + \mathcal{P}^\ell y(t) - y^\ell(t) = \varrho^\ell(t) + \vartheta^\ell(t), \quad \text{for } t \in [0, T] \text{ a.e.}$$

with $\varrho^\ell(t) = y(t) - \mathcal{P}^\ell y(t) \in (V^\ell)^\perp$ and $\vartheta^\ell(t) = \mathcal{P}^\ell y(t) - y^\ell(t) \in V^\ell$. It follows from (5.2) that

$$(15) \quad \int_0^T \|\varrho^\ell(t)\|_X^2 dt = \sum_{i=\ell+1}^{\infty} \lambda_i.$$

Utilizing the definition of the Ritz projection, (2.3) and (5.3) we obtain

$$\begin{aligned} \langle \vartheta_t^\ell(t), \psi \rangle_{V',V} + \langle \vartheta^\ell(t), \psi \rangle_V &= \langle \mathcal{P}^\ell y_t(t) - y_t^\ell(t), \psi \rangle_{V',V} + \langle y(t) - y^\ell(t), \psi \rangle_V \\ &= \langle d(\cdot, y^\ell(t), \mu) - d(\cdot, y(t), \mu), \psi \rangle_H + \langle \mathcal{P}^\ell y_t(t) - y_t(t), \psi \rangle_{V',V} + \langle y(t) - y^\ell(t), \psi \rangle_H. \end{aligned}$$

From the mean value theorem we infer that

$$\begin{aligned} \langle d(\cdot, y^\ell(t), \mu) - d(\cdot, y(t), \mu), \psi \rangle_H &= \int_0^1 \langle d_y(\cdot, \zeta^\ell(t, s), \mu)(y^\ell(t) - y(t)), \psi \rangle_H \\ &= \int_0^1 \int_\Omega \sum_{i=1}^m \mu_i b_i \cosh\left(\zeta^\ell(t, s) \sum_{j=1}^m \mu_j b_j\right) (y^\ell(t) - y(t)) ds \psi d\mathbf{x} \\ &= - \int_0^1 \int_\Omega \sum_{i=1}^m \mu_i b_i \cosh\left(\zeta^\ell(t, s) \sum_{j=1}^m \mu_j b_j\right) (\varrho^\ell(t) + \vartheta^\ell(t)) ds \psi d\mathbf{x} \end{aligned}$$

with $\zeta^\ell(t, s) = y(t) + s(y^\ell(t) - y(t)) \in L^\infty(Q)$ and $s \in [0, 1]$. We define

$$C_1 = \left\| \sum_{i=1}^m \mu_i b_i \right\|_{L^\infty(\Omega)} \max_{s \in [0, T]} \left\| \cosh\left(\zeta^\ell(\cdot, s) \sum_{j=1}^m \mu_j b_j\right) \right\|_{L^\infty(Q)} \geq 0.$$

Then, taking $\psi = \vartheta^\ell(t) \in V^\ell$, using

$$- \int_0^1 \int_\Omega \sum_{i=1}^m \mu_i b_i \cosh\left(\zeta^\ell(t, s) \sum_{j=1}^m \mu_j b_j\right) \vartheta^\ell(t)^2 ds d\mathbf{x} \leq 0$$

and the Young inequality we find that

$$\frac{d}{dt} \|\vartheta^\ell(t)\|_H^2 + \|\vartheta^\ell(t)\|_V^2 \leq C_2 \|\varrho^\ell(t)\|_H^2 + C_3 \|\vartheta^\ell(t)\|_H^2 + \|\mathcal{P}^\ell y_t(t) - y_t(t)\|_{V'}^2,$$

with $C_2 = C_1 + 1$, $C_3 = C_1 + 2$. From the Gronwall lemma and $\vartheta^\ell(0) = \mathcal{P}^\ell y_\circ - y^\ell(0) = 0$ we infer by standard arguments that

$$(16) \quad \int_0^T \|\vartheta^\ell(t)\|_V^2 dt \leq C_4 \left(\sum_{i=\ell+1}^{\infty} \lambda_i + \|\mathcal{P}^\ell y_t(t) - y_t(t)\|_{V'}^2 \right)$$

with a constant $C_4 > 0$. Combining (15) and (16) we conclude that there exists a constant $C > 0$ satisfying (5.4). \square

REFERENCES

- [1] R.A. Adams. *Sobolev Spaces*. Pure and Applied Mathematics, Vol. 65. Academic Press, New York-London, 1975.
- [2] E. Arian, M. Fahl, and E.W. Sachs. Trust-region proper orthogonal decomposition for flow control. Technical Report 2000-25, ICASE, 2000.
- [3] L. Cai and R. E. White. Reduction of model order based on proper orthogonal decomposition for lithium-ion battery simulations. *Journal of The Electrochemical Society*, 156:A154–A161, 2009.
- [4] E. Casas. Pontryagin's principle for state-constrained boundary control problems with finitely many state constraints. *SIAM J. Control and Optimization*, 35:1297–1327, 1997.
- [5] R. Dautray and J.-L. Lions. *Mathematical Analysis and Numerical Methods for Science and Technology. Volume 5: Evolution Problems I*. Springer-Verlag, Berlin, 2000.

- [6] M. Dihlmann and B. Haasdonk. Certified PDE-constrained parameter optimization using reduced basis surrogate models for evolution problems. *Computational Optimization and Applications*, 60:753-787, 2015.
- [7] L.C. Evans. *Partial Differential Equations*. American Math. Society, Providence, Rhode Island, 2002.
- [8] T. Fuller, M. Doyle and J. Newman. Modeling of galvanostatic charge and discharge of the lithium ion battery/polymer/insertion cell. *Journal of the Chemical Society*, 140(6):1526–1533, 1993.
- [9] M. Grepl. Certified reduced basis method for nonaffine linear time-varying and nonlinear parabolic partial differential equations. *WSPC: Mathematical Models and Methods in Applied Sciences*, 22(3):40, 2012.
- [10] M. Gubisch and S. Volkwein. Proper orthogonal decomposition for linear-quadratic optimal control. Accepted, 2015. See <http://kops.ub.uni-konstanz.de/handle/urn:nbn:de:bsz:352-250378>
- [11] M. Hinze, R. Pinnau, M. Ulbrich, and S. Ulbrich. *Optimization with PDE Constraints*. Springer-Verlag, Berlin, 2009.
- [12] M. Hinze and S. Volkwein. Error estimates for abstract linear-quadratic optimal control problems using proper orthogonal decomposition. *Computational Optimization and Applications*, 39:319-345, 2008.
- [13] P. Holmes, J.L. Lumley, G. Berkooz, and C.W. Rowley. *Turbulence, Coherent Structures, Dynamical Systems and Symmetry*. Cambridge Monographs on Mechanics. Cambridge University Press, Cambridge, 2nd ed. edition, 2012.
- [14] L. Iapichino, S. Volkwein, and A. Wesche. A-posteriori error analysis for lithium-ion concentrations in batteries utilizing the reduced basis method. Submitted, 2015. See <http://nbn-resolving.de/urn:nbn:de:bsz:352-0-303692>
- [15] E. Kammann, F. Tröltzsch and S. Volkwein. A method of a-posteriori error estimation with application to proper orthogonal decomposition. *ESAIM: Mathematical Modelling and Numerical Analysis*, 47:555-581, 2013.
- [16] C.T. Kelley. *Iterative methods for optimization*. SIAM Frontiers in Applied Mathematics, no 18, 1999.
- [17] K. Kunisch and S. Volkwein. Galerkin proper orthogonal decomposition methods for parabolic problems. *Numerische Mathematik*, 90:117–148, 2001.
- [18] K. Kunisch and S. Volkwein. Proper orthogonal decomposition for optimality systems. *ESAIM: Mathematical Modelling and Numerical Analysis*, 42:1-23, 2008.
- [19] O. Lass and S. Volkwein. POD Galerkin schemes for nonlinear elliptic-parabolic systems. *SIAM Journal on Scientific Computing*, 35:A1271-A1298, 2013.
- [20] O. Lass and S. Volkwein. Parameter identification for nonlinear elliptic-parabolic systems with application in lithium-ion battery modeling. *Computational Optimization and Applications*, 58:645-677, 2014.
- [21] K. Malanowski, C. Büskens, and H. Maurer. Convergence of approximations to nonlinear control problems. In: Fiacco, A.V. (ed.), *Mathematical Programming with Data Perturbation*, pp. 253-284. Marcel Dekker, New York, 1997.
- [22] Y. Maday, N.C. Nguyen, M. Barrault, and A.T. Patera. An empirical interpolation method: application to efficient reduced-basis discretization of partial differential equations. *Comptes Rendus Mathématique*, 339(9):667672, 2004.
- [23] J. Nocedal and S.J. Wright. *Numerical Optimization*. Springer Series in Operation Research, second edition, 2006.
- [24] J.-P. Raymond and H. Zidani. Hamiltonian Pontryagin’s principles for control problems governed by semilinear parabolic equations. *Applied Mathematics and Optimization*, 39:143-177, 1999.
- [25] L. Sirovich. Turbulence and the dynamics of coherent structures. parts i-ii. *Quarterly of Applied Mathematics*, XVI:561–590, 1987.
- [26] F. Tröltzsch. *Optimal Control of Partial Differential Equations. Theory, Methods and Applications*. American Mathematical Society, Providence, 2010.
- [27] F. Tröltzsch and S. Volkwein. POD a-posteriori error estimates for linear-quadratic optimal control problems. *Computational Optimization and Applications*, 44:83-115, 2009.
- [28] J. Xu, J. Wu and H. Zou. On the well posedness of mathematical model for lithium-ion battery systems. *Methods and Applications of Analysis*, 13:275–298, 2006.

OLIVER LASS, TECHNISCHE UNIVERSITÄT DARMSTADT, FACHBEREICH MATHEMATIK, AG OPTIMIERUNG, DOLIVOSTRASSE 15, 64293 DARMSTADT, GERMANY

E-mail address: `lass@mathematik.tu-darmstadt.de`

STEFAN TRENZ, UNIVERSITÄT KONSTANZ, FACHBEREICH MATHEMATIK UND STATISTIK, UNIVERSITÄTSSTRASSE 10, D-78457 KONSTANZ, GERMANY

E-mail address: `Stefan.Trenz@uni-konstanz.de`

STEFAN VOLKWEIN, UNIVERSITÄT KONSTANZ, FACHBEREICH MATHEMATIK UND STATISTIK, UNIVERSITÄTSSTRASSE 10, D-78457 KONSTANZ, GERMANY

E-mail address: `Stefan.Volkwein@uni-konstanz.de`