A POSTERIORI SNAPSHOT LOCATION FOR POD IN OPTIMAL CONTROL OF LINEAR PARABOLIC EQUATIONS^{\Rightarrow}

ALESSANDRO ALLA^{1,*}, CARMEN GRÄSSLE² AND MICHAEL HINZE²

Abstract. In this paper we study the approximation of an optimal control problem for linear parabolic PDEs with model order reduction based on Proper Orthogonal Decomposition (POD-MOR). POD-MOR is a Galerkin approach where the basis functions are obtained upon information contained in time snapshots of the parabolic PDE related to given input data. In the present work we show that for POD-MOR in optimal control of parabolic equations it is important to have knowledge about the controlled system at the right time instances. We propose to determine the time instances (snapshot locations) by an *a posteriori* error control concept. The proposed method is based on a reformulation of the optimality system of the underlying optimal control problem as a second order in time and fourth order in space elliptic system which is approximated by a space-time finite element method. Finally, we present numerical tests to illustrate our approach and show the effectiveness of the method in comparison to existing approaches.

Mathematics Subject Classification. 49J20, 65N12, 78M34

Received August 29, 2016. Accepted January 15, 2018.

1. INTRODUCTION

Optimization with PDE constraints is nowadays a well-studied topic motivated by its relevance in industrial applications. We are interested in the numerical approximation of such optimization problems in an efficient and reliable way using surrogate models obtained with POD-MOR. The surrogate models are based on *snapshots* from a numerical simulation of the underlying system. Here, snapshots refer to the solution of the system at particular time instances. For the snapshot POD approach we refer the reader to *e.g.* [30]. The knowledge of the snapshots is crucial to build accurate surrogate models. In particular, the accuracy of the POD reduced order model depends on the quality of the chosen snapshots in such a way that they should comply with the physical properties of the underlying system. The aim of this paper is to investigate the computation of well-suited snapshot locations for POD-MOR in optimal control problems governed by linear parabolic equations.

Several works investigated the choice of the snapshots to approximate either dynamical systems or optimal control problems by suitable surrogate models. In [21], the authors proposed to optimize the choice of the time

Keywords and phrases: Optimal control, model order reduction, proper orthogonal decomposition, optimal snapshot location.

 $^{^{\}diamond}$ We would like to thank Z.J. Zhou from Shandong Normal University, China for providing the data and code of the space-time approximation in Gong *et al.* (2012). The first author also acknowledges the support of US Department of Energy (grant number DE-SC0009324).

¹ PUC-Rio, Department of Mathematics, Rua Marques de Sao Vicente, 225, Gavea, 22451-900 Rio de Janeiro, Brazil.

² University of Hamburg, Department of Mathematics, 20146 Hamburg, Germany.

^{*} Corresponding author: alla@mat.puc-rio.br

instances such that the error between POD and the trajectory of the dynamical system is minimized. A recent approach proposes to select the snapshots by an *a posteriori* error estimator in order to equidistribute the state error on the time grid related to the snapshot locations (see *e.g.* [15]). We also mention an adaptive method, proposed in [27], where the aim is to reduce expensive offline costs selecting the snapshots according to an adaptive time-stepping algorithm using time error-control. For further references we refer the interested reader to *e.g.* [27].

In optimal control problems, the computation of the snapshots is not trivial, since state and adjoint state depend on the unknown optimal control. A common approach here is to sample state (and adjoint) snapshots for a reference control, chosen *a priori*, which can be zero or a prediction of the optimal control, for example. Those will constitute the snapshot set on which the construction of the surrogate model is based. However, this approach does not guarantee a meaningful surrogate model since we do not know how well this reference control approximates the sought optimal control. More sophisticated approaches select snapshots by solving an optimization problem in order to improve the selection of the snapshots according to the desired controlled dynamics. For this purpose optimality system POD (OS-POD) is introduced in *e.g.* [20]. In OS-POD, the computation of the POD basis functions is performed by means of the solution of an enlarged optimal control problem which involves the full problem, the reduced equation and the computation of the POD modes. Recently, in [8] different POD basis update strategies are investigated to improve the accuracy of the POD-MOR and avoid the problem of unmodelled dynamics.

The reduction of optimal control problems with particular focus on adaptive adjustments of the surrogate models can be found in *e.g.* [1, 4, 8, 10]. We should also mention another adaptive method for feedback control problems by means of the Hamilton–Jacobi–Bellman equation, introduced in [2].

Recently, an *a posteriori* error estimator was introduced in [16, 32] for optimal control problems. In these works the error between the unknown optimal control and the computed POD suboptimal control is estimated for linear and nonlinear problems, and it is shown that increasing the number of basis functions leads to the desired convergence. OS-POD and *a posteriori* error estimation is combined in [33].

In the present paper we address the question of snapshot location for POD-MOR of optimal control problems. This approach differs from adaptive methods where the basis functions are updated when necessary. To the best of the author's knowledge snapshot location for optimal control problems has not been investigated vet. Our interest focuses on an efficient selection of snapshot locations by means of an *a posteriori* error control approach proposed in [7], which is also generalized to optimal control with control constraints in Theorem 3.1. Our method works as follows: in a first step, we rewrite the optimality conditions as a second order in time and fourth order in space elliptic equation for the adjoint variable. This in particular allows us to apply classical concepts from residual based a *posteriori* error control to construct a suitable time grid for the temporal variable which then serves as snapshot grid for the POD-MOR approximation of the underlying optimal control problem. This is motivated by observations made in [7] related to the temporal *a posteriori* analysis of parabolic optimal control problems, where one outcome was that the structure of the temporal grid is not sensitive against changes in the spatial resolution of the optimal control problem. This motivates us to use the adaptively obtained time grid based on a very coarse spatial resolution as snapshot grid for the POD-MOR approximation of the optimal control problem. We note that the POD model order reduction is done with respect to the spatial variable. Here the novelty for the reduced control problem is twofold: we directly obtain snapshots related to an approximation of the optimal control and, at the same time, we get information about a suitable time grid. We have proposed a similar approach based on a reformulation of the optimality system with respect to the state variable in [3]. Now, we focus our approach on the adjoint variable and generalize the idea presented in [7] to time dependent control intensities with control shape functions including control constraints. Furthermore, we certify our approach by means of several error bounds for the state, adjoint and control variable. We also provide a comparison of the computational cost of our method with OS-POD.

The outline of this paper is as follows. In Section 2 we present the optimal control problem together with the optimality conditions. In Section 3 we recall the main results of [7] and we generalize it to control constraints. Proper orthogonal decomposition and its application to optimal control problems is presented in Section 4. The focus of Section 5 lies in investigating our snapshot location strategy. Finally, numerical tests are discussed in

Section 6 and conclusions are driven in Section 7.

2. Optimal control problem

In this section we describe the optimal control problem. The governing equation is given by a linear parabolic PDE:

where $\Omega \subset \mathbb{R}^q, q \in \{1, 2, 3\}$ is an open bounded domain with smooth boundary, T > 0, $\Omega_T := \Omega \times (0, T]$ is the space-time cylinder, $\Sigma_T := \partial \Omega \times (0, T]$, and the state is denoted by $y : \Omega_T \to \mathbb{R}$. As control space we use $(L^2(0, T; \mathbb{R}^m), \langle \cdot, \cdot \rangle_U)$, where $\langle u, v \rangle_U := \sum_{i=1}^m \langle u_i, v_i \rangle_{L^2(0,T)}$, and define the control operator as $\mathcal{B} : U \to L^2(0, T; H^{-1}(\Omega))$, $(\mathcal{B}u)(t) = \sum_{i=1}^m u_i(t)\chi_i$, where $\chi_i \in H^{-1}(\Omega)(1 \le i \le m)$ denote specified control actions. Thus \mathcal{B} is linear and bounded. For the control variable we require

$$u \in U_{ad} := \{ u \in U \mid u_a(t) \le u(t) \le u_b(t) \text{ in } \mathbb{R}^m \text{ a.e. in } [0,T] \} \subset L^{\infty}(0,T;\mathbb{R}^m)$$

with $u_a, u_b \in L^{\infty}(0, T; \mathbb{R}^m), u_a(t) \leq u_b(t)$ almost everywhere in (0, T). It is well-known (see *e.g.* [22]) that for a given initial condition $y_0 \in L^2(\Omega)$ and a forcing term $f \in L^2(0, T; H^{-1}(\Omega))$ equation (2.1) admits a unique solution $y = y(u) \in W(0, T)$, where

$$W(0,T) := \left\{ v \in L^2\left(0,T; H^1_0(\Omega)\right), \frac{\partial v}{\partial t} \in L^2\left(0,T; H^{-1}(\Omega)\right) \right\}.$$

If $y_0 \in H_0^1(\Omega)$, higher regularity results can be derived according to [5].

The weak formulation of (2.1) is given by: find $y \in W(0,T)$ with $y(0) = y_0$ and

$$\int_{0}^{T} \left(\langle y_t(t), v \rangle_{H^{-1}, H^1_0} + \int_{\Omega} \nabla y(x, t) \nabla v(x, t) \mathrm{d}x \right) \mathrm{d}t = \int_{0}^{T} \langle f(t) + \mathcal{B}u(t), v \rangle_{H^{-1}, H^1_0} \mathrm{d}t \quad \text{ for all } v \in L^2(0, T; H^1_0(\Omega)).$$

$$(2.2)$$

The cost functional we want to minimize is given by

$$J(y,u) := \frac{1}{2} \|y - y_d\|_{L^2(\Omega_T)}^2 + \frac{\alpha}{2} \|u\|_U^2,$$
(2.3)

where $y_d \in L^2(\Omega_T)$ is the desired state and the regularization parameter α is a real positive constant. The optimal control problem then reads

$$\min_{u \in U_{ad}} \hat{J}(u) := J(y(u), u), \text{ where } y(u) \text{ satisfies (2.1)}.$$
(2.4)

Note that U_{ad} is a non-empty, bounded, convex and closed subset of $L^{\infty}(0,T;\mathbb{R}^m)$. Hence, it is easy to argue that (2.4) admits a unique solution $u \in U_{ad}$ with associated state $y(u) \in W(0,T)$, see *e.g.* [22].

A. ALLA ET AL.

The first order optimality system of the optimal control problem (2.4) is given by the state equation (2.1), together with the adjoint equation

$$\begin{array}{rcl} -p_t - \Delta p &=& y - y_d & \text{in } \Omega_T, \\ p(\cdot, T) &=& 0 & \text{in } \Omega, \\ p &=& 0 & \text{on } \Sigma_T, \end{array} \right\}$$

$$(2.5)$$

and the variational inequality

$$\langle \alpha u + \mathcal{B}^* p, v - u \rangle_U \ge 0 \quad \text{for all } v \in U_{ad},$$

$$(2.6)$$

where $\mathcal{B}^* : L^2(0,T; H^{-1}(\Omega))^* \to U^*$ is the dual operator of \mathcal{B} . In (2.6) we have identified $L^2(0,T; H^{-1}(\Omega))^*$ with $L^2(0,T; H_0^1(\Omega))$ and U^* with U, where we use that Hilbert spaces are reflexive. The variational inequality (2.6) is equivalent to the projection formula

$$u(t) = \mathcal{P}_{U_{ad}} \left\{ -\frac{1}{\alpha} (\mathcal{B}^* p)(t) \right\} \text{ for almost all } t \in [0, T],$$
(2.7)

where $\mathcal{P}_{U_{ad}}: U \to U_{ad}$ denotes the orthogonal projection onto U_{ad} . It follows from the reflexivity of the involved spaces that the action of the adjoint operator \mathcal{B}^* is given as

$$(\mathcal{B}^*v)(t) = \left(\langle \chi_1, v \rangle_{H^{-1}, H^1_0}, \dots, \langle \chi_m, v \rangle_{H^{-1}, H^1_0} \right)$$

and

$$\mathcal{P}_{U_{ad}}\left\{-\frac{1}{\alpha}\mathcal{B}^*p\right\}_i = \max\left\{u_a, \min\{u_b, -\frac{1}{\alpha}\langle\chi_i, p\rangle_{H^{-1}, H_0^1}\}\right\}.$$

Since our domain is smooth, the regularity of the optimal state, the optimal control and the associated adjoint state are limited through the regularity of the initial state y_0 , the right hand side f, the control u and the desired state y_d .

3. Space-time approximation

In this section, we consider the reformulation of the optimality system (2.1)-(2.5)-(2.6) as an elliptic equation of fourth order in space and second order in time for the adjoint variable p. We refer to [7, 24] for more details. Following these works, we include control constraints. Here, we aim to derive an *a posteriori* error estimate for the time discretization as suggested in [7], which then turns out to be the basis for our model reduction approach to solve (2.4).

We define

$$H_0^{2,1}(\Omega_T) := \left\{ v \in H^{2,1}(\Omega_T) : v(T) = 0 \text{ in } \Omega \right\},\$$

where

$$H^{2,1}(\Omega_T) = L^2(0,T; H^2(\Omega)) \cap H^1_0(\Omega)) \cap H^1(0,T; L^2(\Omega))$$

is equipped with the norm

$$\|w\|_{H^{2,1}(\Omega_T)}^2 := \left(\|w\|_{L^2(0,T;H^2(\Omega))}^2 + \|w\|_{H^1(0,T;L^2(\Omega))}^2\right).$$

Under the assumptions $y_0 \in H_0^1(\Omega)$, $\chi_i \in L^2(\Omega)$ for i = 1, ..., m and $y_d \in H^{2,1}(\Omega_T)$, the regularity of $y, p \in H^{2,1}(\Omega_T)$ is ensured, see [5] for the details. Then, the first order optimality conditions (2.1)–(2.5)–(2.6) can be transformed into an initial boundary value problem for p in space-time:

$$-p_{tt} + \Delta^2 p - \mathcal{BP}_{U_{ad}} \left(-\frac{1}{\alpha} \mathcal{B}^* p \right) = -(y_d)_t + \Delta y_d \quad \text{in } \Omega_T,$$

$$p(\cdot, T) = 0 \qquad \text{in } \Omega,$$

$$p = 0 \qquad \text{on } \Sigma_T,$$

$$\Delta p = y_d \qquad \text{on } \Sigma_T,$$

$$(p_t + \Delta p)(0) = y_d(0) - y_0 \qquad \text{in } \Omega,$$

$$(3.1)$$

where, without loss of generality, we have set $f \equiv 0$. We note that the quantity

$$\mathcal{BP}_{U_{ad}}\left(-\frac{1}{\alpha}\mathcal{B}^*p\right)$$

is nondifferentiable and nonlinear in p and thus (3.1) becomes a semilinear second order in time and fourth order in space elliptic problem with a monotone nonlinearity. Existence of a unique weak solution for (3.1) can be proved analogously to [24] and follows from the fact that the optimal control problem (2.4) in the case of control constraints with closed and convex $U_{ad} \subset U$ admits a unique solution.

In order to provide the weak formulation of (3.1), we define the operator A_0 and the linear form L_0 as

$$A_0: H_0^{2,1}(\Omega_T) \times H_0^{2,1}(\Omega_T) \to \mathbb{R}, \qquad L_0: H_0^{2,1}(\Omega_T) \to \mathbb{R}$$

$$A_0(v,w) := \int_{\Omega_T} \left(v_t w_t - \mathcal{BP}_{U_{ad}} \left(-\frac{1}{\alpha} \mathcal{B}^* v \right) w \right) + \int_{\Omega_T} \Delta v \Delta w + \int_{\Omega} \nabla v(0) \cdot \nabla w(0),$$

$$L_0(v) := \int_0^T \langle -\frac{\partial y_d}{\partial t} + \Delta y_d, v \rangle_{H^{-1}(\Omega) \times H^1_0(\Omega)} - \int_\Omega (y_d(0) - y_0)v(0) + \int_{\Sigma_T} y_d \nabla v \cdot \hat{n},$$

where \hat{n} denotes the outer normal to the boundary $\partial\Omega$. The weak formulation of equation (3.1) for given $y_d \in H^{2,1}(\Omega_T), y_0 \in H^1_0(\Omega)$, reads:

find
$$p \in H_0^{2,1}(\Omega_T)$$
 with $A_0(p,v) = L_0(v) \quad \forall v \in H_0^{2,1}(\Omega_T).$ (3.2)

It follows from the monotonicity of the orthogonal projection that (3.2) admits a unique solution p, compare e.g. ([14], Thm. 1.25). We put our attention on the semi-discrete approximation of (3.1) and investigate a priori and a posteriori error estimates for the time discrete problem, where the space is kept continuous. Let us consider the time discretization $0 = t_0 < t_1 < \cdots < t_n = T$ with $\Delta t_j = t_j - t_{j-1}$ and $\Delta t := \max_j \Delta t_j$. Let $I_j := [t_{j-1}, t_j]$. We define the time discrete space

$$V_t^k := \left\{ v \in H^{2,1}(\Omega_T) : \ v(\cdot)|_{I_j} \in P_1(I_j) \right\}, \qquad \bar{V}_t^k := V_t^k \cap H_0^{2,1}(\Omega_T),$$

where the notation $P_1(I_j)$ stands for the polynomials of first order on the interval I_j and k refers to the time discretization. Then, we consider the semi-discrete problem:

find
$$p_k \in \bar{V}_t^k$$
 with $A_0(p_k, v_k) = L_0(v_k), \quad \forall v_k \in \bar{V}_t^k.$ (3.3)

Using the arguments of e.g. ([14], Thm. 1.25) one can show that problem (3.3) admits a unique solution $p_k \in \bar{V}_t^k$. We note that with (3.2) and (3.3) we have the Galerkin orthogonality

$$A_0(p, v_k) - A_0(p_k, v_k) = 0 \quad \forall v_k \in \bar{V}_t^k.$$
(3.4)

Thus, for $v \in H_0^{2,1}(\Omega_T)$ it holds true

$$A_0(p,v) - A_0(p_k,v) = A_0(p,v-v_k) - A_0(p_k,v-v_k), \qquad \forall v_k \in \bar{V}_t^k.$$

The following theorem states a temporal residual type a posteriori error estimate for p, which generalizes the estimation of ([7], Thm. 3.5) to the control constrained optimal control problem (2.4):

Theorem 3.1. Let $p \in H_0^{2,1}(\Omega_T)$ and $p_k \in \overline{V}_t^k$ denote the solutions to (3.2) and (3.3), respectively. Then we obtain

$$\|p - p_k\|_{H^{2,1}(\Omega_T)}^2 \le C_1 \eta^2, \tag{3.5}$$

where $C_1 > 0$ and

$$\eta^2 = \sum_j \Delta t_j^2 \int_{I_j} \left\| -\frac{\partial y_d}{\partial t} + \Delta y_d + \frac{\partial^2 p_k}{\partial t^2} + \mathcal{BP}_{U_{ad}} \left(-\frac{1}{\alpha} \mathcal{B}^* p_k \right) - \Delta^2 p_k \right\|_{L^2(\Omega)}^2 + \sum_j \int_{I_j} \|y_d - \Delta p_k\|_{L^2(\partial\Omega)}^2.$$

Proof. We start the proof showing a consequence of the monotonicity of the projector operator $-\mathcal{P}_{U_{ad}}\{-\frac{1}{\alpha}\mathcal{B}^*p\}$. We find that

$$\left\langle -\mathcal{P}_{U_{ad}}\left\{-\frac{1}{\alpha}\mathcal{B}^*p_1\right\} + \mathcal{P}_{U_{ad}}\left\{-\frac{1}{\alpha}\mathcal{B}^*p_2\right\}, \mathcal{B}^*p_1 - \mathcal{B}^*p_2\right\rangle_U \ge 0, \quad \forall p_1, p_2 \in H_0^{2,1}(\Omega_T),$$

and hence

$$\int_{\Omega_T} \left(-\mathcal{BP}_{U_{ad}} \left\{ -\frac{1}{\alpha} \mathcal{B}^* p_1 \right\} + \mathcal{BP}_{U_{ad}} \left\{ -\frac{1}{\alpha} \mathcal{B}^* p_2 \right\} \right) (p_1 - p_2) \ge 0.$$
(3.6)

For easier notation, we set $N(p) := -\mathcal{BP}_{U_{ad}} \left\{ -\frac{1}{\alpha} \mathcal{B}^* p \right\}$. Let $e^p := p - p_k$ and let $\pi_k e^p$ denote the standard Lagrange type temporal interpolation of e^p . Using the inequality

$$\|v\|_{H^{2,1}(\Omega_T)}^2 \le C\left(\|\frac{\partial v}{\partial t}\|_{L^2(\Omega_T)}^2 + \|\Delta v\|_{L^2(\Omega_T)}^2\right)$$

for $v \in H_0^{2,1}(\Omega_T)$ and C > 0 from ([7], Lem. 2.5), the monotonicity (3.6) and the Galerkin orthogonality (3.4), we can estimate:

$$\begin{split} c\|p-p_k\|_{H^{2,1}(\Omega_T)}^2 \\ &\leq \left\|\frac{\partial(p-p_k)}{\partial t}\right\|_{L^2(\Omega_T)}^2 + \|\Delta(p-p_k)\|_{L^2(\Omega_T)}^2 \\ &\leq \left\|\frac{\partial(p-p_k)}{\partial t}\right\|_{L^2(\Omega_T)}^2 + \|\Delta(p-p_k)\|_{L^2(\Omega_T)}^2 + \int_{\Omega_T} (N(p)-N(p_k))(p-p_k) \\ &= \int_{\Omega_T} \frac{\partial(p-p_k)}{\partial t} \frac{\partial e^p}{\partial t} + \int_{\Omega_T} \Delta(p-p_k)\Delta e^p + \int_{\Omega_T} (N(p)-N(p_k))e^p \\ &= \int_{\Omega_T} \frac{\partial(p-p_k)}{\partial t} \frac{\partial(e^p-\pi_k e^p)}{\partial t} + \int_{\Omega_T} \Delta(p-p_k)\Delta(e^p-\pi_k e^p) + \int_{\Omega_T} (N(p)-N(p_k))(e^p-\pi_k e^p) \\ &= \int_{\Omega_T} (-\frac{\partial y_d}{\partial t} + \Delta y_d)(e^p-\pi_k e^p) + \int_{\Sigma_T} y_d \nabla(e^p-\pi_k e^p) \cdot \hat{n} - \int_{\Omega_T} \frac{\partial p_k}{\partial t} \frac{\partial(e^p-\pi_k e^p)}{\partial t} \\ &- \int_{\Omega_T} \Delta p_k \Delta(e^p-\pi_k e^p) - \int_{\Omega_T} N(p_k)(e^p-\pi_k e^p) \end{split}$$

Integration by parts on each time interval and Green's formula lead to

$$c\|p - p_k\|_{H^{2,1}(\Omega_T)}^2$$

$$\leq \sum_j \int_{I_j} \int_{\Omega} \left(-\frac{\partial y_d}{\partial t} + \Delta y_d + \frac{\partial^2 p_k}{\partial t^2} - \Delta^2 p_k - N(p_k) \right) (e^p - \pi_k e^p) + \sum_j \int_{I_j} \int_{\partial\Omega} (y_d - \Delta p_k) \nabla (e^p - \pi_k e^p) \cdot \hat{n}.$$

Utilizing error estimates of the Lagrange interpolation π_k , the trace inequality and Young's inequality, we find

$$\begin{split} \|p - p_k\|_{H^{2,1}(\Omega_T)}^2 &\leq C_1 \sum_j \Delta t_j^2 \int_{I_j} \left\| -\frac{\partial y_d}{\partial t} + \Delta y_d + \frac{\partial^2 p_k}{\partial t^2} - \Delta^2 p_k + \mathcal{BP}_{U_{ad}} \left\{ -\frac{1}{\alpha} \mathcal{B}^* p_k \right\} \right\|_{L^2(\Omega)}^2 \\ &+ C_1 \sum_j \int_{I_j} \|y_d - \Delta p_k\|_{L^2(\partial\Omega)}^2. \end{split}$$

Theorem 3.1 provides a tool to refine the time grid by means of the residual of the system (3.1). Due to (2.7), the time instances of this grid may be regarded as ideal snapshot locations for POD-MOR applied to problem (2.4).

4. POD FOR OPTIMAL CONTROL PROBLEMS

In this section, we recall the POD method which we use to replace the original problem (2.4) by a surrogate model. The main interest when applying the POD method is to reduce computation times and storage capacity while retaining a satisfying approximation quality. This is possible due to the key fact that POD basis functions (unlike typical finite element ansatz functions) contain information about the underlying model, since the POD modes are derived from snapshots of a solution data set. Usually, we are able to improve the accuracy of a POD suboptimal solution by enlarging the number of utilized POD basis functions or enriching the snapshot set, for instance. The snapshot form of POD proposed by Sirovich in [30] works in the continuous version as follows.

A. ALLA ET AL.

Let us suppose that the continuous solution y(t) of (2.1) and p(t) of (2.5) belong to a real separable Hilbert space V, where $V = H_0^1(\Omega)$ or $L^2(\Omega)$, equipped with its inner product $\langle \cdot, \cdot \rangle$ and associated norm $\|\cdot\|^2 = \langle \cdot, \cdot \rangle$. We set $\mathcal{V} := \operatorname{span}\{z^k(t) \mid t \in [0,T] \text{ and } 1 \leq k \leq 3\} \subseteq V$, where $z^1(t) := y(t), z^2(t) := p(t), z^3(t) := \dot{p}(t)$. Note that the initial condition $y(0) = y_0$ is included in \mathcal{V} . The aim is to determine a POD basis $\{\psi_1, \ldots, \psi_\ell\} \subset V$ of rank $\ell \in \{1, \ldots, d\}$ with $d = \dim(\mathcal{V}) \leq \infty$, by solving the following constrained minimization problem:

$$\min_{\psi_1,\dots,\psi_\ell} \sum_{k=1}^3 \int_0^T \left\| z^k(t) - \sum_{i=1}^\ell \langle z^k(t), \psi_i \rangle \psi_i \right\|^2 \mathrm{d}t \quad \text{s.t. } \langle \psi_j, \psi_i \rangle = \delta_{ij} \quad \text{for } 1 \le i, j \le \ell,$$
(4.1)

where δ_{ij} denotes the Kronecker symbol, *i.e.* $\delta_{ij} = 0$ for $i \neq j$ and $\delta_{ii} = 1$.

It is well-known (see [9]) that a solution to problem (4.1) is given by the first ℓ eigenvectors $\{\psi_1, \ldots, \psi_\ell\}$ corresponding to the ℓ largest eigenvalues $\lambda_i > 0$ of the self-adjoint linear operator $\mathcal{R} : V \to V$, *i.e.* $\mathcal{R}\psi_i = \lambda_i\psi_i$, $i = 1, \ldots, \ell$, where \mathcal{R} is defined as follows:

$$\mathcal{R}\psi = \sum_{k=1}^{3} \int_{0}^{T} \langle z^{k}(t), \psi \rangle \ z^{k}(t) dt \quad \text{for } \psi \in V.$$

Moreover, we can quantify the POD approximation error by the neglected eigenvalues (more details in [9]) as follows:

$$\sum_{k=1}^{3} \int_{0}^{T} \left\| z^{k}(t) - \sum_{i=1}^{\ell} \langle z^{k}(t), \psi_{i} \rangle \psi_{i} \right\|^{2} \mathrm{d}t = \sum_{i=\ell+1}^{d} \lambda_{i}.$$
(4.2)

Let us assume that we have computed POD basis functions $\{\psi_i\}_{i=1}^{\ell}$. Then, we define the POD Galerkin ansatz of order ℓ for the state y as:

$$y^{\ell}(t) = \sum_{i=1}^{\ell} w_i(t)\psi_i,$$
(4.3)

where $y^{\ell} \in V^{\ell} := \operatorname{span}\{\psi_1, \ldots, \psi_{\ell}\}$ and the unknown coefficients are denoted by $\{w_i\}_{i=1}^{\ell}$. If we plug this ansatz into the weak formulation of the state equation (2.2) and use V^{ℓ} as the test space, we get the following reduced order model for (2.2) of low dimension:

$$\int_{0}^{T} \left(\langle y_{t}^{\ell}(t), \psi \rangle_{H^{-1}, H_{0}^{1}} + \int_{\Omega} \nabla y^{\ell}(t, x) \cdot \nabla \psi(x) dx \right) dt = \int_{0}^{T} \langle (f + \mathcal{B}u)(t), \psi \rangle_{H^{-1}, H_{0}^{1}} dt \\
\forall \psi \in V^{\ell} \text{ and } t \in (0, T] \text{ a.e.}, \qquad (4.4)$$

$$\int_{\Omega} y^{\ell}(0) \psi dx = \int_{\Omega} y_{0} \psi dx$$

Choosing $\psi = \psi_i$ for $i = 1, ..., \ell$ and utilizing (4.3), we infer from (4.4) that the coefficients $(w_1(t), \ldots, w_\ell(t)) =: w(t)$ satisfy

$$M^{\ell} \dot{w}(t) + A^{\ell} w(t) = F^{\ell}(t) \quad \text{a.e. in } (0,T], \quad M^{\ell} w(0) = y_0^{\ell},$$

where $(M^{\ell})_{ij} = \int_{\Omega} \psi_j \psi_i dx$, $(A^{\ell})_{ij} = \int_{\Omega} \nabla \psi_j \cdot \nabla \psi_i dx$, $(F^{\ell}(t))_j = \int_0^T (f + \mathcal{B}u)(t)\psi_j dx$ and $(y_0^{\ell})_j = \int_{\Omega} y_0 \psi_j dx$. Note that M^{ℓ} is the identity matrix, if we choose as inner product $\langle \cdot, \cdot \rangle := \langle \cdot, \cdot \rangle_{L^2(\Omega)}$.

The reduced order model surrogate (ROM) for the optimal control problem is given by

$$\min_{u \in U_{ad}} \hat{J}^{\ell}(u) \text{ s.t. } y^{\ell}(u) \text{ satisfies (4.4)},$$
(4.5)

where \hat{J}^{ℓ} is the reduced cost functional, *i.e.* $\hat{J}^{\ell}(u) := \hat{J}(y^{\ell}(u), u)$. We recall that the discretization of the optimal solution \bar{u}^{ℓ} to (4.5) is determined by the relation between the adjoint state and control and refer to [11] for more details about the variational discretization concept.

In order to solve the reduced optimal control problem (4.5), we consider the well-known first order optimality condition given by the variational inequality

$$\langle \nabla \hat{J}^{\ell}(\bar{u}^{\ell}), u - \bar{u}^{\ell} \rangle_U \ge 0 \quad \forall u \in U_{ad},$$

which is sufficient since the underlying problem is convex.

By construction of the POD space the first order optimality conditions of (4.5) also deliver that the adjoint POD scheme for the approximation of p is given by: find $p^{\ell}(t) \in V^{\ell}$ with $p^{\ell}(T) = 0$ satisfying

5. SNAPSHOT LOCATION STRATEGY FOR OPTIMAL CONTROL PROBLEM

In Section 4, we recalled the POD method in the infinite dimensional settings, where the POD basis functions are computed in such a way that the error between the state y(t) and adjoint p(t) trajectories and its POD projection is minimized in (4.1). In practice, we can only compute approximate solutions $\{y(t_j)\}_{j=0}^n$ and

 ${p(t_j)}_{j=0}^n$ for some given time instances $0 = t_0 < t_1 < \cdots < t_n = T$. Hence, we introduce the following time-discrete version of (4.1):

$$\min_{\psi_1,\dots,\psi_\ell} \sum_{k=1}^3 \sum_{j=0}^n \beta_j \left\| z^k(t_j) - \sum_{i=1}^\ell \langle z^k(t_j), \psi_i \rangle \psi_i \right\|^2, \text{ s.t. } \langle \psi_j, \psi_i \rangle = \delta_{ij} \quad \text{ for } 1 \le i,j \le \ell$$

$$(5.1)$$

which is obtained by applying quadrature to the time integral in (4.1) using the trapezoidal rule with weights β_0, \ldots, β_n .

We note that problem (5.1) constitutes a strong dependence of the POD basis functions on the chosen snapshot locations t_0, \ldots, t_n . The related snapshots shall have the property to capture the main features of the dynamics of the truth solution as much as possible. Here it is important to select suitable time instances at which important features of the optimal solution are located. A natural question is:

How to pick time instances that represent good locations for snapshots in POD-MOR for (4.5)?

This problem is usually addressed in the offline stage for POD, which is the phase needed for snapshot generation, POD basis computation and the projection of the dynamical system into the low-dimensional space. Nevertheless, the computation of the snapshots in optimal control problems is not straightforward due to the presence of the unknown control u(t) in the dynamical system. Snapshots are usually computed from a given input control $u_0 \in U_{ad}$ (a very common choice is $u_0 \equiv 0$), since we do not have information about the problem. It turns out that this approach might be inappropriate and therefore the POD basis functions need to be updated as suggested in e.g. [8]. These circumstances raise the question about the quality of the POD basis and of the

POD suboptimal solution. Our approach focuses on the snapshot location for optimal control problems. For this reason we take advantage of the *a posteriori* error estimator (3.5) in Section 3 which provides a suitable location of time instances for the POD adjoint state and at the same time we get an approximation of the optimal control which can be used as an input control u_0 in order to generate the snapshots.

The use in the offline-stage of equation (3.1) allows to overcome the choice of an input control u_0 and to select snapshot locations according to the *a posteriori* error estimation presented in Theorem 3.1. We note that the ellipticity of equation (3.1) plays a crucial role in this approach. The same approach would not work, if one solves the optimality conditions directly. The numerical approximation of *p* provides important information about the control input. In fact, due to the variational inequality (2.6), we are first able to build an approximate control *u* and finally compute the associated state y(u). In this way our snapshot set will contain information about the state corresponding to an approximation of the optimal control. Thanks to this numerical approximation of the optimal control problem we can build the snapshot matrix and compute the POD basis functions where the number ℓ is chosen such that $\sum_{i=\ell+1}^{d} \lambda_i \approx 0$.

To summarize, the approximation of equation (3.1) is very useful in model order reduction since we overcome the choice of the initial input control to generate the snapshot set. Moreover, we also gain information about a temporal grid, which allows us to better resolve p with respect to time.

The *a posteriori* error estimation (3.5) guarantees that the finite element approximation of (3.1) in the time variable is below a certain tolerance. Therefore, the reduced optimal control problem (4.5) is set up and solved on the resulting adaptive time grid. Now the question is:

How good is the quality of the computed time grid in terms of the error between the optimal solution and the POD surrogate solution?

5.1. Error analysis for the adjoint variable

Let us motivate our approach by analyzing the error $\|p(u) - \tilde{p}_k^{\ell}(u_k^{\ell})\|_{L^2(0,T,V)}$ between the optimal adjoint solution p(u) of (2.5) associated with the optimal control u for (2.4), *i.e.* $u = \mathcal{P}_{U_{ad}}(-\frac{1}{\alpha}\mathcal{B}^*p)$ and the POD reduced approximation $\tilde{p}_k^{\ell}(u_k^{\ell})$, which is the time discrete solution to the POD-ROM for (2.5) associated with the time discrete optimal control u_k^{ℓ} for (4.5), *i.e.* $y = y(u_k^{\ell})$ in (2.5). We denote by V the space $V = H_0^1(\Omega)$ and by H the space $L^2(\Omega)$.

To ease notation let us denote by $p_k(u_k)$ the time discrete adjoint solution of (3.3) associated with the control $u_k = \mathcal{P}_{U_{ad}}(-\frac{1}{\alpha}\mathcal{B}^*p_k)$, and by $\tilde{p}_k(u_k)$ the time discrete adjoint solution to (2.5) with respect to the control u_k . Furthermore, $p_k(u_k^\ell)$ is the time discrete adjoint solution to (2.5) with respect to the suboptimal control u_k^ℓ , *i.e.* $y = y(u_k^\ell)$ in (2.5). By $\mathcal{P}^\ell : V \to V^\ell$ we denote the orthogonal POD projection operator as follows:

$$\mathcal{P}^{\ell} y := \sum_{i=1}^{\ell} \langle y, \psi_i \rangle_V \psi_i \quad \text{ for } y \in V$$

Proposition 5.1. Suppose that $\ell > 0$, p(u) is the solution of (2.5) and $\tilde{p}_k^{\ell}(u_k^{\ell})$ is the time discrete solution of (2.5). Let us also assume that

$$\|p_k(u_k) - \tilde{p}_k(u_k)\|_{L^2(0,T;V)} \le \varepsilon.$$
(5.2)

Then, there exist $C_1, C_2, C_3 > 0$ such that

$$\|p(u) - \tilde{p}_{k}^{\ell}(u_{k}^{\ell})\|_{L^{2}(0,T,V)} \leq \sqrt{C_{1}}\eta + \frac{C_{2}}{\alpha}(\|\zeta_{k}\|_{U} + \|\zeta_{k}^{\ell}\|_{U}) + \sqrt{C_{3}\left(\sum_{i=\ell+1}^{d} \lambda_{i}^{k} + \|y_{k} - y_{k}^{\ell}\|_{L^{2}(0,T,H)}^{2}\right)}.$$
 (5.3)

Proof. By the triangular inequality we get the following estimates for the $L^2(0,T;V)$ -norm:

$$\|p(u) - \tilde{p}_{k}^{\ell}(u_{k}^{\ell})\| \leq \underbrace{\|p(u) - p_{k}(u_{k})\|}_{(5.4.1)} + \underbrace{\|p_{k}(u_{k}) - \mathcal{P}^{\ell}p_{k}(u_{k})\|}_{(5.4.2)} + \underbrace{\|\mathcal{P}^{\ell}p_{k}(u_{k}) - \mathcal{P}^{\ell}\tilde{p}_{k}(u_{k}^{\ell})\|}_{(5.4.3)} + \underbrace{\|\mathcal{P}^{\ell}\tilde{p}_{k}(u_{k}^{\ell}) - \tilde{p}_{k}^{\ell}(u_{k}^{\ell})\|}_{(5.4.4)}$$

$$(5.4)$$

The term (5.4.1) can be estimated by (3.5) and concerns the snapshot generation. Thus, we can decide on a certain tolerance in order to have a prescribed error. The second term (5.4.2) in (5.4) is the POD projection error and can be estimated by the sum of the neglected eigenvalues. Then, we note that the third term (5.4.3)can be estimated as follows:

$$\|\mathcal{P}^{\ell}p_{k}(u_{k}) - \mathcal{P}^{\ell}\tilde{p}_{k}(u_{k}^{\ell})\| \leq \|\mathcal{P}^{\ell}\| \|p_{k}(u_{k}) - \tilde{p}_{k}(u_{k}^{\ell})\| \leq C_{2}\|u_{k} - u_{k}^{\ell}\|_{U},$$
(5.5)

where $\|\mathcal{P}^{\ell}\| \leq 1$ and $C_2 > 0$ is the constant referring to the Lipschitz continuity of p_k independent of k as in [23]. In equation (5.5) we make use of assumption (5.2).

In order to control the quantity $||u_k - u_k^{\ell}||_U \leq ||u_k - u||_U + ||u - u_k^{\ell}||_U$ we make use of the *a posteriori* error estimation of [32], which provides an upper bound for the error between the (unknown) optimal control and any arbitrary control u_p (here $u_p = u_k$ and $u_p = u_k^{\ell}$) by

$$\|u-u_p\|_U \le \frac{1}{\alpha} \|\zeta_p\|_U,$$

where α is the regularization parameter in the cost functional and $\zeta_p \in L^2(0,T;\mathbb{R}^m)$ is chosen such that

$$\langle \alpha u_p - \mathcal{B}^* p(u_p) + \zeta_p, u - u_p \rangle_U \ge 0 \quad \forall u \in U_{ad}$$

is satisfied. Finally, the last term (5.4.4) can be estimated according to [13] and involves the sum of the eigenvalues not considered, the first derivative of the time discrete adjoint variable and the difference between the state and the POD state:

$$\|\mathcal{P}^{\ell}p_{k}(u_{k}^{\ell}) - p_{k}^{\ell}(u_{k}^{\ell})\|^{2} \leq C_{3}\left(\sum_{i=\ell+1}^{d}\lambda_{i}^{k} + \|\dot{p}_{k}(u_{k}^{\ell}) - \mathcal{P}^{\ell}\dot{p}_{k}(u_{k}^{\ell})\|_{L^{2}(0,T,V')}^{2} + \|y_{k}(u_{k}^{\ell}) - y_{k}^{\ell}(u_{k}^{\ell})\|_{L^{2}(0,T,H)}^{2}\right), \quad (5.6)$$

for a constant $C_3 > 0$. We note that the sum of the neglected eigenvalues is sufficiently small provided that ℓ is large enough. Furthermore, the error estimation (5.6) depends on the time derivative \dot{p}_k . To avoid this dependence, we include time derivative information concerning the adjoint variable into the snapshot set, see [19].

Remark 5.1.

- 1. We assume in (5.2) that the difference between $p_k(u_k)$ and $\tilde{p}_k(u_k)$ is small since the continuous solution of (3.3) coincides with solution of (2.5).
- 2. We note that estimations (5.3) and (5.6) involve the state variable which is estimated in the following Section 5.2.

5.2. Error analysis for the state variable

In this section we address the problem of the certification of the quality for POD approximation for the state variable. It may happen that the time grid selected for the adjoint p will not be accurate enough for the state variable y. Therefore a further refinement of the time grid might be useful in order to reduce the error between the POD state and the true state below a given threshold. This is not guaranteed if we use the time

A. ALLA ET AL.

grid, which results from the use of the estimate (3.5). Here, we consider the error between the full solution $y(u_k^{\ell})$ corresponding to the suboptimal control u_k^{ℓ} and the time discrete POD solution $y_k^{\ell}(u_k^{\ell})$, where we assume to have the same temporal grid for snapshots and the solution of our POD reduced order problem. In this situation, the following estimate is proved in [19]:

$$\sum_{j=0}^{n} \beta_{j} \|y(t_{j}; u_{k}^{\ell}) - y_{j}^{\ell}(u_{k}^{\ell})\|_{H}^{2} \leq \sum_{j=1}^{n} \left(\Delta t_{j}^{2} C_{y}((1+c_{p}^{2})\|y_{tt}(u_{k}^{\ell})\|_{L^{2}(I_{j},H)}^{2} + \|y_{t}(u_{k}^{\ell})\|_{L^{2}(I_{j};V)}) \right)$$
(5.7a)

$$+\sum_{j=1}^{n} C_y \left(\sum_{i=\ell+1}^{d} \left(|\langle \psi_i, y_0 \rangle_V|^2 + \lambda_i \right) \right)$$
(5.7b)

$$+\sum_{j=1}^{n}\sum_{i=\ell+1}^{d}C_{y}\frac{\lambda_{i}}{\Delta t_{j}^{2}}$$
(5.7c)

where $C_y > 0$ is a constant depending on T, but independent of the time grid $\{t_j\}_{j=0}^n$. We note that $y(t_j; u_k^\ell)$ is the continuous solution of (2.1) at given time instances related to the suboptimal control u_k^ℓ . The temporal step size in the subinterval $[t_{j-1}, t_j]$ is denoted by Δt_j . The positive weights β_j are given by

$$\beta_0 = \frac{\Delta t_1}{2}, \quad \beta_j = \frac{\Delta t_j + \Delta t_{j+1}}{2} \text{ for } j = 1, \dots, n-1, \text{ and } \beta_n = \frac{\Delta t_n}{2}.$$

The constant c_p is an upper bound of the projection operator. A similar estimate can be carried out for the V-norm. We refer the interested reader to [19].

Estimate (5.7) provides now a recipe for further refinement of the time grid in order to approximate the state y within a prescribed tolerance. One option here consists in equidistributing the error contributions of the term (5.7a), while the number of modes has to be adapted to the time grid size according to the term (5.7c). Finally, the number ℓ of modes should be chosen such that the term in (5.7b) remains within the prescribed tolerance.

5.3. The algorithm

The *a posteriori* error control concept for (3.1) now offers the possibility to select snapshot locations by a time adaptive procedure. For this purpose, (3.1) is solved adaptively in time, where the spatial resolution (Δx in Algorithm 5.1) is chosen to be very coarse in order to keep the computational costs low. This is possible due to the fact that spatial and temporal discretization decouple when using the solution technique of [7] as we will see in Section 6, compare Figure 3. The resulting time grid points now serve as snapshot locations, on which our POD reduced order model for the optimization is based. The snapshots are now obtained from a simulation on the time adaptive grid of (2.1) and (2.5) with fine resolution h. Then, we generate time derivative adjoint snapshots by finite differences. The right-hand side u in the simulation of (2.1) is obtained from (2.6) with p from (2.5) computed with spatially coarse resolution Δx . The certification of the state variable is then performed according to (5.7) as a post-processing procedure. Finally, we set up and solve the POD-ROM problem (4.5) where the POD model reduction is performed with respect to the spatial dimension This strategy might not deliver the optimal time instances, but it is a practical and efficient strategy, which turns out to deliver good approximation results (compare Sect. 6) at low costs.

The algorithm is summarized below in Algorithm 5.1.

Algorithm 5.1. Adaptive snapshot selection for optimal control problems.

Require: coarse spatial grid size Δx , fine spatial grid size h, maximal number of degrees of freedom (dof) for the adaptive time discretization, T > 0.

1: Solve (3.1) adaptively w.r.t. time with spatial resolution Δx and obtain the time grid \mathcal{T} with solution $p_{\Delta x}$.

2: Set $u_{\Delta x} = \mathcal{P}_{U_{ad}}\left(-\frac{1}{\alpha}\mathcal{B}^*p_{\Delta x}\right).$

3: Solve (2.1) on \mathcal{T} with spatial resolution Δx corresponding to the control $u_{\Delta x}$.

4: Refine the time interval \mathcal{T} according to (5.7) and construct the time grid \mathcal{T}_{new} .

5: Generate state and adjoint snapshots by solving (2.1) with r.h.s. $u_{\Delta x}$ and (2.5), respectively, on \mathcal{T}_{new} with spatial resolution h. Generate time derivative adjoint snapshots with time finite differences on those adjoint snapshots.

6: Compute a POD basis of order ℓ and build the POD reduced order model (4.5) based on the state, adjoint state and time derivative adjoint state snapshots.

adjoint state and time derivative adjoint state snapshot

7: Solve (4.5) with the time grid \mathcal{T}_{new}

6. Numerical tests

In our numerical computations we use a one-dimensional spatial domain and a finite element discretization in space by means of conformal piecewise linear polynomials. In order to treat (3.1) numerically, we use a space-time mixed finite element method analogue to [7]. We generate the snapshots and the POD-ROM (4.5) by using the implicit Euler method in time and piecewise linear and continuous finite elements in space. The solution of the optimal control problem (4.5) is combined with a projected gradient method with stopping criteria $\|\hat{J}'(u^k)\| \leq \tau_r \|\hat{J}'(u^k)\|_U + \tau_a$ and an Armijo linesearch. We note that τ_r, τ_a are the relative and absolute tolerance, respectively. In the following numerical examples, we apply Algorithm 5.1 in order to validate this strategy by numerical results.

The numerical tests illustrate that a time adaptive grid for snapshot location and for POD reduced order optimal control delivers more accurate approximation results than a uniform time grid. We show three different numerical tests. The first example presents a steep gradient at the end of the time interval in the adjoint variable. In the second example the adjoint state develops an interior layer in the middle of the time interval and finally we introduce control contraints in the third example. Moreover we also show the benefits of the post processing for the state variable (step 4 in Algorithm 5.1) to achieve more accurate approximation results for both state and adjoint state.

All coding is done in MATLAB R2015a and the computations are performed on a 2.50 GHz computer with 8 Gb RAM. To solve the nonlinear, nonsmooth equation (3.1), we have extended the MATLAB codes of [7] to control constrained systems and computed a solution with a fixed point iteration.

6.1. Test 1: solution with steep gradient towards final time

The data for this test example is inspired from Example 5.3 in [7], with the following choices: $\Omega = (0, 1)$ and [0, T] = [0, 1]. We set $U_{ad} = L^{\infty}(0, T; \mathbb{R}^m)$. The example is built such that the exact optimal solution (\bar{y}, \bar{u}) of problem (2.4) with associated optimal adjoint state \bar{p} is known:

$$\bar{y}(x,t) = \sin(\pi x)\sin(\pi t),$$

$$\bar{p}(x,t) = x(x-1)\left(t - \frac{e^{(t-1)/\varepsilon} - e^{-1/\varepsilon}}{1 - e^{-1/\varepsilon}}\right),$$

$$\bar{u}(t) = -\frac{1}{\alpha}\mathcal{B}^*\bar{p}(x,t) = -t + \frac{e^{(t-1)/\varepsilon} - e^{-1/\varepsilon}}{1 - e^{-1/\varepsilon}}$$
(6.1)

A. ALLA ET AL.



FIGURE 1. Test 1: analytical optimal adjoint state \bar{p} (*left*), POD adjoint solution p^{ℓ} utilizing an equidistant time grid with $\Delta t = 1/20$ (*middle*), POD adjoint solution p^{ℓ} utilizing an adaptive time grid with dof = 21 (*right*).



FIGURE 2. Test 1: contour lines of the analytical optimal adjoint state \bar{p} (*left*), POD adjoint solution p^{ℓ} utilizing an equidistant time grid with $\Delta t = 1/20$ (*middle*), POD adjoint solution p^{ℓ} utilizing an adaptive time grid with dof = 21 (*right*).

with m = 1 and the control shape function $\chi(x) = x(x-1)$ for the operator \mathcal{B} . This leads to the right hand side

$$f(x,t) = \pi \sin(\pi x)(\cos(\pi t) + \pi \sin(\pi t)) + x(x-1)\left(t - \frac{e^{(t-1)/\varepsilon} - e^{-1/\varepsilon}}{1 - e^{-1/\varepsilon}}\right)$$

the desired state

$$y_d(x,t) = \sin(\pi x)\sin(\pi t) + x(x-1)\left(1 - \frac{e^{(t-1)/\varepsilon} \cdot 1/\varepsilon}{1 - e^{-1/\varepsilon}}\right) + 2\left(t - \frac{e^{(t-1)/\varepsilon} - e^{-1/\varepsilon}}{1 - e^{-1/\varepsilon}}\right)$$

and the initial condition $y_0(x) = 0$. We choose the regularization parameter to be $\alpha = 1/30$. For small values of ε (we use $\varepsilon = 10^{-4}$), the adjoint state \bar{p} develops a layer towards t = 1, which can be seen in the left plots of Figures 1 and 2.

In this test run we focus on the influence of the time grid to approximate of the POD solution. Therefore, we compare the use of two different types of time grids: an equidistant time grid characterized by the time increment $\Delta t = 1/n$ and a non-equidistant (adaptive) time grid characterized by n + 1 degrees of freedom (dof). We build the POD-ROM from the uncontrolled problem; we create the snapshot ensemble by determining the associated state $y(u_0)$ and adjoint state $p(u_0)$ corresponding to the control function $u_0 \equiv 0$ and we also include the initial condition y_0 and the time derivatives of the adjoint $p_t(u_0)$ into our snapshot set, which is accomplished with time finite differences of the adjoint snapshots. We use $\ell = 1$ POD basis function. Although we would also



FIGURE 3. Test 1: adaptive space-time grids with dof = 21 according to the strategy in [7] and $\Delta x = 1/100$ (*left*) and $\Delta x = 1/5$ (*middle*), respectively, and the equidistant grid with $\Delta t = 1/20$ (*right*).



FIGURE 4. Test 1: analytical optimal control \bar{u} (top left), approximation $u_{\Delta x}$ of the optimal control gained by step 1 of Algorithm 5.1 (top middle), approximation of the optimal control utilizing OS-POD on a uniform time grid with $\Delta t = 1/20$ (top right); POD control utilizing a uniform time grid with $\Delta t = 1/20$ (bottom left), POD control utilizing an adaptive time grid with dof = 21 (bottom middle), approximation of the optimal control utilizing OS-POD on an adaptive time grid with dof = 21 (bottom right).

have the possibility to use suboptimal snapshots corresponding to an approximation $u_{\Delta x}$ of the optimal control, here, we want to emphasize the importance of the time grid. Nevertheless in this example, the quality of the POD solution does not really differ, if we consider suboptimal or uncontrolled snapshots. First, we leave out the post-processing step 4 of Algorithm 5.1 and discuss the inclusion of it later.

Figure 3 visualizes the space-time mesh of the numerical solution of (3.1) with the temporal residual type a posteriori error estimate (3.5). The first grid in Figure 3 corresponds to the choice of dof = 21 and $\Delta x = 1/100$, whereas the grid in the middle refers to using dof = 21 and $\Delta x = 1/5$. Both choices for spatial discretization

-							
Δt	$\varepsilon_{ m abs}^y$	$\varepsilon^u_{ m abs}$	$\varepsilon^p_{ m abs}$	dof	$\varepsilon_{ m abs}^{y}$	$\varepsilon^u_{ m abs}$	$\varepsilon^p_{ m abs}$
1/20	$1.5120 \cdot 10^{-02}$	$1.9837 \cdot 10^{-01}$	$3.6247 \cdot 10^{-02}$	21	$5.1874 \cdot 10^{-02}$	$5.3428 \cdot 10^{-02}$	$9.6343 \cdot 10^{-03}$
1/42	$1.1186 \cdot 10^{-02}$	$2.1071 \cdot 10^{-01}$	$3.8490 \cdot 10^{-02}$	43	$5.1634 \cdot 10^{-02}$	$2.4868 \cdot 10^{-02}$	$4.3611 \cdot 10^{-03}$
1/61	$1.0774 \cdot 10^{-02}$	$2.1447 \cdot 10^{-01}$	$3.9173 \cdot 10^{-02}$	62	$5.1599 \cdot 10^{-02}$	$2.3275 \cdot 10^{-02}$	$4.0691 \cdot 10^{-03}$
1/114	$1.1157 \cdot 10^{-02}$	$2.1846 \cdot 10^{-01}$	$3.9893 \cdot 10^{-02}$	115	$5.1568 \cdot 10^{-02}$	$2.3027 \cdot 10^{-02}$	$4.0340 \cdot 10^{-03}$
1/6500	$2.0485 \cdot 10^{-02}$	$1.2893 \cdot 10^{-01}$	$2.3480 \cdot 10^{-02}$	_	_	_	_

TABLE 1. Test 1: absolute errors between the analytical optimal solution and the POD solution depending on the time discretization (equidistant: columns 1–4, adaptive: columns 5–8).

TABLE 2. Test 1: CPU times for POD offline computations (snapshot generation and POD basis computation) depending on the time discretization (equidistant: columns 1–3, adaptive: columns 4–7.

Δt	POD offline	POD online	dof	Compute \mathcal{T}	POD offline	POD online
1/20	$0.12\mathrm{s}$	$0.04\mathrm{s}$	21	$2.3\mathrm{s}$	$0.12\mathrm{s}$	$0.04\mathrm{s}$
1/42	$0.13\mathrm{s}$	$0.07\mathrm{s}$	43	$4.2\mathrm{s}$	$0.13\mathrm{s}$	$0.07\mathrm{s}$
1/61	$0.14\mathrm{s}$	$0.09\mathrm{s}$	62	$5.5\mathrm{s}$	$0.14\mathrm{s}$	$0.09\mathrm{s}$
1/114	$0.18\mathrm{s}$	$0.10\mathrm{s}$	115	$8.9\mathrm{s}$	$0.18\mathrm{s}$	$0.10\mathrm{s}$
1/6500	$58.6\mathrm{s}$	$4.74\mathrm{s}$	_	_	_	_

TABLE 3. Absolute errors between the analytical optimal solution and the finite element solution of spatial resolution h = 1/100 depending on the resolution of the uniform time discretization.

Δt	$\varepsilon^y_{ m abs}$	$\varepsilon^u_{ m abs}$	$\varepsilon^p_{ m abs}$	J
1/20	$1.2961 \cdot 10^{-02}$	$1.9898 \cdot 10^{-01}$	$3.6325 \cdot 10^{-02}$	$4.1652 \cdot 10^{+04}$
1/42	$6.8850 \cdot 10^{-03}$	$2.1144 \cdot 10^{-01}$	$3.8602 \cdot 10^{-02}$	$1.9834 \cdot 10^{+04}$
1/61	$5.1979 \cdot 10^{-03}$	$2.1528 \cdot 10^{-01}$	$3.9303 \cdot 10^{-02}$	$1.3656 \cdot 10^{+04}$
1/114	$3.5501 \cdot 10^{-03}$	$2.1939 \cdot 10^{-01}$	$4.0054 \cdot 10^{-02}$	$7.3078 \cdot 10^{+03}$
1/6500	$1.2343 \cdot 10^{-03}$	$1.3016 \cdot 10^{-01}$	$2.3716 \cdot 10^{-02}$	$1.4116 \cdot 10^{+02}$
1/20000	$5.0444 \cdot 10^{-04}$	$5.2073 \cdot 10^{-02}$	$9.3888 \cdot 10^{-03}$	$9.0788 \cdot 10^{+01}$
1/40000	$2.8080 \cdot 10^{-04}$	$2.7981 \cdot 10^{-02}$	$4.9088 \cdot 10^{-03}$	$8.5692 \cdot 10^{+01}$

lead to the exact same time grid, which displays fine time steps towards the end of the time horizon (where the layer in the optimal adjoint state is located), whereas at the beginning and in the middle of the time interval the time steps are larger. This clearly indicates that the resulting time adaptive grid is very insensitive against changes in the spatial resolution. For the sake of completeness, the equidistant grid with the same number of degrees of freedom is shown in the right plot of Figure 3.

Since the generation of the time adaptive grid as well as the approximation of the optimal solution is done in the offline computation part of POD-MOR, this process shall be performed quickly, which is why we pick $\Delta x = 1/5$ for step 1 in Algorithm 5.1.

In the middle and right panel of Figures 1 and 2, we show the surface and contour lines of the POD adjoint state with an equidistant and adaptive time grid, respectively. The analytical control intensity $\bar{u}(t)$, the approximation $u_{\Delta x}$ of the optimal control computed in step 1 of Algorithm 5.1 as well as the POD controls utilizing a uniform and time adaptive grid, respectively, are shown in Figure 4.

Table 1 summarizes the approximation quality of the POD solution depending on different time discretizations. The fineness of the time discretization (characterized by Δt and dof, respectively) is chosen in such a way that the results of uniform and adaptive temporal discretization are comparable. The absolute errors

TABLE 4. Computational times of the full finite element solution compared with the computational times of the POD-MOR solution including all offline times.

Δt	h	Full FE run	dof	Compute \mathcal{T}	POD offline	POD online	Speedup
$1/6500 \\ 1/20000 \\ 1/40000$	$1/100 \\ 1/100 \\ 1/100$	$6.15 \\ 16.81 \\ 34.50$	21	2.3	0.12	0.04	$2.5 \\ 6.8 \\ 14.0$
$1/6500 \\ 1/20000 \\ 1/40000$	$1/1000 \\ 1/1000 \\ 1/1000$	$37.51 \\ 115.62 \\ 231.38$	21	2.3	0.16	0.04	$15.0 \\ 46.2 \\ 92.5$

TABLE 5. Test 1: evaluation of each summand of the error estimation (5.3).

dof	$\varepsilon^p_{ m abs}$	η_p^i	η_p^b	$\ \zeta_k\ _U + \ \zeta_k^\ell\ _U$	$\sum_{i=\ell+1}^d \lambda_i$
21	$9.6343 \cdot 10^{-03}$	$4.9518 \cdot 10^{+00}$	$4.8031 \cdot 10^{-04}$	$1.6033 \cdot 10^{-02}$	$3.3938 \cdot 10^{-04}$
43	$4.3611 \cdot 10^{-03}$	$1.1976 \cdot 10^{+00}$	$5.0087 \cdot 10^{-05}$	$1.9200 \cdot 10^{-02}$	$2.9454 \cdot 10^{-04}$
62	$4.0691 \cdot 10^{-03}$	$7.2852 \cdot 10^{-01}$	$2.9835 \cdot 10^{-05}$	$1.9707 \cdot 10^{-02}$	$2.9212 \cdot 10^{-04}$
115	$4.0340 \cdot 10^{-03}$	$3.4966 \cdot 10^{-01}$	$1.4845 \cdot 10^{-05}$	$2.0191 \cdot 10^{-02}$	$2.9090 \cdot 10^{-04}$

TABLE 6. Test 1: value of the cost functional at the POD solution utilizing uniform and adaptive time discretization, respectively, analytical value: $J \approx 8.3988 \cdot 10^{+01}$.

Δt	$J(y^\ell, u)$	dof	$J(y^\ell, u)$
1/20	$4.1652 \cdot 10^{+04}$	21	$8.7960 \cdot 10^{+01}$
1/42	$1.9834 \cdot 10^{+04}$	43	$8.4252 \cdot 10^{+01}$
1/61	$1.3656 \cdot 10^{+04}$	62	$8.4102 \cdot 10^{+01}$
1/114	$7.3078 \cdot 10^{+03}$	115	$8.4034 \cdot 10^{+01}$
1/6500	$1.4116 \cdot 10^{+02}$	_	_

between the analytical optimal state \bar{y} and the POD solution y^{ℓ} , defined by $\varepsilon_{abs}^{y} := \|\bar{y} - y^{\ell}\|_{L^{2}(\Omega_{T})}$, are listed in columns 2 and 6; same applies for the errors in the control $\varepsilon_{abs}^{u} := \|\bar{u} - u^{\ell}\|_{\mathcal{U}}$ (columns 3 and 7) and adjoint state $\varepsilon_{abs}^{p} := \|\bar{p} - p^{\ell}\|_{L^{2}(\Omega_{T})}$ (columns 4 and 8). If we compare the results, we note that we gain one order of accuracy for the adjoint and control variable with the time adaptive grid. For the state variable, the use of an adaptive time grid leads to slightly worse results, which we will discuss later. Both for the full solution and the reduced order solution, three steepest descent iterations are needed. In order to achieve an accuracy in the control variable of order 10^{-2} and an accuracy in the adjoint state of order 10^{-3} with an equidistant time grid, we would need about $n = 20\,000$ time steps (not listed in Tab. 1) which is not feasible with our machine. The largest number of time steps we are able to manage with our computing resources is n = 6500. With n = 6500 time steps, the accuracy in the control and adjoint state is one order worse than the results obtained with the time adaptive grid with only 21 degrees of freedom. Furthermore, the offline CPU time for snapshot generation and POD basis computation is 58.6 s. On the contrary, Algorithm 5.1 leads to a total CPU time of 2.46 s if we consider offline and online altogether (compare Tab. 2). This emphasizes that using an appropriate (non-equidistant) time grid for the adjoint variable is of particular importance in order to efficiently achieve POD controls of good quality.

Furthermore, we compare the optimization of the full FE model to the optimization with the POD-MOR model obtained with our approach. First of all, Table 3 lists the errors between the FE solution with spatial



FIGURE 5. Test 1: analytical optimal state \bar{y} (top left), desired state y_d (top right); POD state y^{ℓ} utilizing a uniform time grid with $\Delta t = 1/20$ (bottom left), POD state y^{ℓ} utilizing an adaptive time grid with dof = 21 (bottom right).

TABLE 7. Test 1: improvement of approximation quality concerning the state variable and the corresponding CPU times. The initial time grid \mathcal{T} is computed with dof = 43.

$N_{\rm refine}$	$\varepsilon^y_{ m abs}$	$\varepsilon^u_{ m abs}$	$\varepsilon^p_{ m abs}$	CPU time
0	$5.1874 \cdot 10^{-02}$	$5.3428 \cdot 10^{-02}$	$9.6343 \cdot 10^{-03}$	_
5	$4.0058 \cdot 10^{-02}$	$2.1145 \cdot 10^{-02}$	$3.6378 \cdot 10^{-03}$	$0.2\mathrm{s}$
10	$3.0909 \cdot 10^{-02}$	$1.8396 \cdot 10^{-02}$	$3.0895 \cdot 10^{-03}$	$0.3\mathrm{s}$
20	$2.4759 \cdot 10^{-02}$	$1.7104 \cdot 10^{-02}$	$2.8210 \cdot 10^{-03}$	$0.4\mathrm{s}$
30	$2.3028 \cdot 10^{-02}$	$1.6971 \cdot 10^{-02}$	$2.7906 \cdot 10^{-03}$	$0.4\mathrm{s}$

discretization h = 1/100 and the analytical solution. It shows that we need more about $n = 40\,000$ equidistantly distributed time steps to achieve a value of the cost functional close to the analytical value which is $J \approx 8.3988 \cdot 10^{+01}$. Our POD-MOR approach with an adaptive time grid reaches this value with only dof = 115 (see Tab. 6).

We note that we were only able to realize the computation for $n \ge 20\,000$ on a compute server for memory reasons. On top of that, we compare in Table 4 the computational times of the full FE optimization using a uniform time grid with the POD-MOR solution utilizing our approach. If we want to achieve a FE solution with

TABLE 8. Test 1: accuracy of the approximate control $u_{\Delta x}$ from Algorithm 5.1 in comparison with the approximate OS-POD control u_{OSPOD} on uniform (columns 2 and 4) and adaptive (column 3) time grids.

error	$\Delta t = 1/20$	dof = 21	$\Delta t = 1/6500$
$\begin{aligned} \ u_{OSPOD} - \bar{u}\ _U \\ \ u_{\Delta x} - \bar{u}\ _U \end{aligned}$	$2.2988 \cdot 10^{-01}$	$\begin{array}{c} 9.2801 \cdot 10^{-02} \\ 1.1162 \cdot 10^{-02} \end{array}$	$1.2896 \cdot 10^{-01}$

TABLE 9. Test 1: computational times for computing an approximation of the optimal control using step 1–2 of Algorithm 5.1 (row 2) and using OS-POD for different temporal resolutions (rows 3–4).

Control	CPU time
$u_{\Delta x}, \operatorname{dof} = 21$	$2.3 { m s}$
$u_{OSPOD}, \Delta t = 1/20$	$0.37 { m s}$
$u_{OSPOD}, \Delta t = 1/6500$	$315.6 { m s}$

similar accuracy as the POD solution on a time adaptive grid, a very large number of time steps are needed, which leads to high computational times, so that we get large speedup factors. Finally, we want to mention that solving (2.4) with the adaptive space-time approach of [7], it takes 43.8 s for dof = 21 and h = 1/100 and 2111.5 s for dof = 21 and h = 1/1000 (compare Tab. 2). We note that our method has a speedup factor of 17.8 and 851, respectively.

Table 5 contains the evaluations of each term in (5.3). The value η_p^i (η_p^b) refers to the first (second) part in (3.5). For this test example, we note that the term η_p^i influences the estimation. However, we observe that the better the semi-discrete adjoint state $p_{\Delta x}$ from step 1 of Algorithm 5.1 is, the better will be the POD adjoint solution. Since all summands of (5.3) can be estimated, Table 5 allows us to control the approximation of the POD adjoint state. The estimation (5.7) concerning the state variable will be investigated later on.

Furthermore, a comparison of the value of the cost functional is given in Table 6. The aim of the optimization problem (2.4) is to minimize the quantity of interest J(y, u). The analytical value of the cost functional at the optimal solution is $J(\bar{y}, \bar{u}) \approx 8.3988 \cdot 10^1$. Table 6 clearly points out that the use of a time adaptive grid is fundamental for solving the optimal control problem (2.4). The huge differences in the values of the cost functional is due to the great increase of the desired state y_d at the end of the time interval (see Fig. 5). Small time steps at the end of the time interval, as it is the case in the time adaptive grid, lead to much more accurate results.

Now, let us discuss the inclusion of step 4 in Algorithm 5.1. Since we went for an adaptive time grid regarding the adjoint variable, we cannot in general expect that the resulting time grid is a good time grid for the state variable. Table 1 confirms that a uniform time grid leads to better approximation results in the state variable than the time adaptive grid. In order to improve also the approximation quality in the state variable, we incorporate the error estimation (5.7) from [19] in a post-processing step after producing the time grid with the strategy of [7] and before starting the POD solution process. We define

$$\eta_{\text{POD}_{j}} := \Delta t_{j}^{2} \left(\int_{I_{j}} (\|y_{tt}^{k}\|_{H}^{2} + \|y_{t}^{k}\|_{V}^{2}) \right)$$

where $y_t^k \approx y_t(t_k)$ and $y_{tt}^k \approx y_{tt}(t_k)$ are computed *via* finite difference approximation. We perform bisection on those time intervals I_j , where the quantity η_{POD_j} has its maximum value and repeat this N_{refine} times. This results in the time grid \mathcal{T}_{new} . The improvement in the approximation quality in the state variable can be seen in Table 7. The more additional time instances we include according to (5.7), the better the approximation results get with respect to the state. Moreover, also the approximation quality in the control and adjoint state

is improved. The CPU times for this post-processing step are listed in Table 7. We note that the sum of the neglected eigenvalues $\sum_{i=2}^{d} \lambda_i$ is approximately zero and the second largest eigenvalue of the correlation matrix is of order 10^{-10} , which makes the use of additional POD basis functions redundant. Likewise, in this particular example the choice of richer snapshots (even the optimal snapshots) does not bring significant improvements in the approximation quality of the POD solutions. So, this example shows that solely the use of an appropriate adaptive time mesh efficiently improves the accuracy of the POD solution.

Finally, we compare the approximation $u_{\Delta x}$ for the optimal control from step 2 in Algorithm 5.1 with an approximation of the optimal control we get by performing OS-POD¹ (optimality system POD, see e.g. [20]) on a uniform time grid. In our runs for OS-POD, the snapshots are taken from the state, adjoint state, time derivative of the adjoint state and the initial condition y_0 . We use $\ell = 1$ basis function and perform two gradient steps. The comparison of the controls $u_{\Delta x}$ (Fig. 4, top middle) and u_{OSPOD} on a uniform time grid (Fig. 4, top right) with the optimal control \bar{u} (Fig. 4, top left) visualizes that $u_{\Delta x}$ is closer to the optimal solution than u_{OSPOD} . We also combined OS-POD with the time adaptive grid (Fig. 4, bottom right). In this example, it turns out that the accuracy of the control variable is improved by a well-suited adaptive time grid. Tables 8 and 9 show the control error and the CPU time. As expected, a very large number of time steps for a uniform time discretization is needed. Indeed, the control computed on a fine equidistant grid with n = 6500 is less accurate than a corse adaptive grid (compare Tab. 8). Finally, Table 9 shows the computational costs of the offline stage to compute a reference control with OS-POD and our approach.

6.2. Test 2: solution with steep gradient in the middle of the time interval

Let $\Omega = (0,1)$ be the spatial domain and [0,T] = [0,1] be the time interval. We choose $\varepsilon = 10^{-4}$ and $\alpha = 1$. To begin with, we consider an unconstrained optimal control problem and investigate the inclusion of control constraints separately in Test 3. We build the example in such a way that the analytical solution (\bar{y}, \bar{u}) of (2.4) is given by:

$$\bar{y}(x,t) = x^3(x-1)t, \quad \bar{p}(x,t) = \sin(\pi x)\operatorname{atan}\left(\frac{t-0.5}{\varepsilon}\right)(t-1),$$

$$\bar{u}_1(t) = \bar{u}_2(t) = -\operatorname{atan}\left(\frac{t-0.5}{\varepsilon}\right)(t-1)\left(\frac{32}{\pi^3} - \frac{8}{\pi^2}\right),$$

$$\bar{\chi}_1(x) = \max\left\{0, 1 - 16(x - 0.25)^2\right\}, \quad \bar{\chi}_2(x) = \max\left\{0, 1 - 16(x - 0.75)^2\right\}.$$

The desired state and the forcing term are chosen accordingly. Due to the arcus-tangent term and the small value for ε , the adjoint state exhibits an interior layer with steep gradient at t = 0.5, which can be seen in the left panel of Figures 6 and 7. The shape functions χ_1 and χ_2 are shown in Figure 8 on the left side. As in Test 1, we study the use of two different time grids: an equidistant time discretization and the time adaptive grid computed in step 1 of Algorithm 5.1 (see Fig. 9). Once again, we note that spatial and temporal discretization decouple when computing the time adaptive grid utilizing the *a posteriori* estimation (3.5), which enables us to use a large spatial resolution Δx for solving the elliptic system and to keep the offline costs low.

We choose state and adjoint snapshots as well as time derivative adjoint snapshots corresponding to $u_0 = 0$ and we also include the initial condition y_0 into our snapshot set. We take $\ell = 4$ POD modes. Later we will also

¹OS-POD concerns the computation of the POD basis functions in optimal control problems. Therefore, the POD problem (4.1) is included in the whole optimization process, *i.e.* the POD reduced order model is computed from the snapshots corresponding to the optimal control. Thus, the POD basis are optimal for the investigated problem. However, the method might turn out to be computationally as expensive as the original optimal control problem, and for practical realizations, the idea is to update the POD basis in the direction of the minimum. We refer the interested reader to [20] for a complete description of the method.



FIGURE 6. Test 2: analytical optimal adjoint state \bar{p} (*left*), POD adjoint solution p^{ℓ} with $\ell = 4$ utilizing an equidistant time grid with $\Delta t = 1/40$ (*middle*), POD adjoint solution p^{ℓ} with $\ell = 4$ utilizing an adaptive time grid with dof = 41 (*right*).



FIGURE 7. Test 2: contour lines of the analytical optimal adjoint state \bar{p} (*left*), POD adjoint solution p^{ℓ} with $\ell = 4$ utilizing an equidistant time grid with $\Delta t = 1/40$ (*middle*), POD adjoint solution p^{ℓ} with $\ell = 4$ utilizing an adaptive time grid with dof = 41 (*right*).



FIGURE 8. Test 2: shape functions $\chi_1(x)$ and $\chi_2(x)$ (*left*), decay of the eigenvalues on semilog scale (*middle*) and first POD basis function ψ_1 (*right*) utilizing uniform time grid with $\Delta t = 1/40$.

try out different numbers of utilized POD basis functions. The middle and right plots of Figures 6 and 7 show the surface and contour lines of the POD adjoint solution utilizing an equidistant time grid (with $\Delta t = 1/40$) and utilizing the adaptive time grid (with dof = 41), respectively. Clearly, the equidistant time grid fails to capture the interior layer at t = 1/2 satisfactorily, whereas the POD adjoint state utilizing the adaptive time grid approximates the interior layer accurately.



FIGURE 9. Test 2: adaptive space-time grids with dof = 41 according to the strategy in [7] and $\Delta x = 1/100$ (*left*) and $\Delta x = 1/5$ (*middle*), respectively, and the equidistant grid with $\Delta t = 1/40$ (*right*).



FIGURE 10. Test 6.2: analytical optimal state \bar{y} (*left*), POD solution y^{ℓ} with $\ell = 4$ utilizing an equidistant time grid with $\Delta t = 1/40$ (*middle*), POD solution y^{ℓ} with $\ell = 4$ utilizing an adaptive time grid with dof = 41 (*right*).



FIGURE 11. Test 6.2: contour lines of the analytical optimal state \bar{y} (*left*), POD solution y^{ℓ} with $\ell = 4$ utilizing an equidistant time grid with $\Delta t = 1/40$ (*middle*), POD solution y^{ℓ} with $\ell = 4$ utilizing an adaptive time grid with dof = 41 (*right*).

Unlike Test Example 6.1, the adaptive time grid is also a suitable time grid for the state variable in this numerical test example. This can be seen visually when comparing the results for the POD state utilizing uniform discretization and utilizing the adaptive time grid with the analytical optimal state, Figures 10 and 11.

Table 10 summarizes the absolute errors between the analytical optimal solution and the POD solution for the state, control and adjoint state for all test runs with an equidistant and adaptive time grid, respectively. If we compare the results of the numerical approximation, we note that the use of an adaptive time grid heavily

TABLE 10. Test 6.2: absolute errors between the analytical optimal solution and the POD solution with $\ell = 4$ depending on the time discretization (equidistant: columns 1–4, adaptive: columns 5–8).

Δt	$\varepsilon^y_{ m abs}$	$\varepsilon^u_{ m abs}$	$\varepsilon^p_{ m abs}$	dof	$\varepsilon^y_{ m abs}$	$\varepsilon^u_{ m abs}$	$\varepsilon_{\rm abs}^p$
1/20	$5.0767 \cdot 10^{-01}$	$7.8419 \cdot 10^{+00}$	$3.5413 \cdot 10^{+01}$	21	$4.0346 \cdot 10^{-02}$	$5.4053 \cdot 10^{-01}$	$2.4409 \cdot 10^{+00}$
1/40	$2.6242 \cdot 10^{-01}$	$4.1058 \cdot 10^{+00}$	$1.8542 \cdot 10^{+01}$	41	$2.2178 \cdot 10^{-04}$	$5.3471 \cdot 10^{-03}$	$1.3186 \cdot 10^{-02}$
1/68	$1.5603 \cdot 10^{-01}$	$2.4503 \cdot 10^{+00}$	$1.1065 \cdot 10^{+01}$	69	$9.7031 \cdot 10^{-05}$	$4.5702 \cdot 10^{-03}$	$4.2670 \cdot 10^{-03}$
1/134	$7.8741 \cdot 10^{-02}$	$1.2386 \cdot 10^{+00}$	$5.5938 \cdot 10^{+00}$	135	$8.5577 \cdot 10^{-05}$	$4.4901 \cdot 10^{-03}$	$2.3507 \cdot 10^{-03}$
1/6500	$1.4104 \cdot 10^{-04}$	$4.9177 \cdot 10^{-03}$	$9.3324 \cdot 10^{-03}$	_	_	_	_

TABLE 11. Test 2: CPU times for POD offline computations (snapshot generation and POD basis computation) depending on the time discretization (equidistant: columns 1–3, adaptive: columns 4–7).

Δt	POD offline	POD online	dof	Compute \mathcal{T}	POD offline	POD online
1/20	$0.16\mathrm{s}$	$0.07\mathrm{s}$	21	$1.89\mathrm{s}$	$0.16\mathrm{s}$	$0.07\mathrm{s}$
1/40	$0.17\mathrm{s}$	$0.08\mathrm{s}$	41	$5.32\mathrm{s}$	$0.17\mathrm{s}$	$0.08\mathrm{s}$
1/68	$0.17\mathrm{s}$	$0.10\mathrm{s}$	69	$7.81\mathrm{s}$	$0.17\mathrm{s}$	$0.10\mathrm{s}$
1/134	$0.19\mathrm{s}$	$0.17\mathrm{s}$	135	$12.71\mathrm{s}$	$0.19\mathrm{s}$	$0.17\mathrm{s}$
1/6500	$52.4\mathrm{s}$	$5.99\mathrm{s}$	_	_	_	_

TABLE 12. Test 2: value of the cost functional with $\ell = 4$, true value $J \approx 1.0085 \cdot 10^{+03}$.

Δt	$J(y^\ell, u)$	dof	$J(y^\ell, u)$
1/20	$3.1225 \cdot 10^{+05}$	21	$1.9553 \cdot 10^{+04}$
1/40	$1.5619 \cdot 10^{+05}$	41	$1.0274 \cdot 10^{+03}$
1/68	$9.1901 \cdot 10^{+04}$	69	$1.0065 \cdot 10^{+03}$
1/134	$4.6655 \cdot 10^{+04}$	135	$1.0082 \cdot 10^{+03}$
1/10000	$1.0350 \cdot 10^{+03}$	—	_

improves the quality of the POD solution with respect to an equidistant grid. In fact, we get an improvement of order four.

The exact optimal control intensities $\bar{u}_1(t)$ and $\bar{u}_2(t)$ as well as the POD solutions utilizing uniform and adaptive temporal discretization are illustrated in Figure 12.

Another point of comparison is the evaluation of the cost functional. Since the exact optimal solution is known analytically, we can compute the exact value of the cost functional, which is $J(\bar{y}, \bar{u}) = 1.0085 \cdot 10^3$. As expected, the adaptive time grid enables us to approximate this value of the cost functional quite well when using dof = 135, see Table 12. In contrast, the use of a very fine temporal discretization with $\Delta t = 1/10000$ is still worse than the results with the adaptive time grid with only 41 degrees of freedom. Again, this emphasizes the importance of a suitable time grid.

For computing the full solution and the reduced order solution, we need three gradient steps in each case. The CPU times for the test runs are summarized in Table 11. In order to achieve an accuracy in the control and adjoint variable of order 10^{-3} , we need around n = 6500 time steps. In this case, the CPU time for the POD offline phase gets really large (52.4 s). In contrast, computing a time adaptive grid on which the snapshots are sampled and the POD-ROM simulation is performed, makes computationally sense.

For the sake of completeness we also study and compare the POD approximation for $\ell = 1$ POD basis function. To begin, we note that the decay of the eigenvalues are in the middle of Figure 8. As one can see, the



FIGURE 12. Test 2: analytical control intensities $\bar{u}_1(t)$ (top left) and $\bar{u}_2(t)$ (bottom left), POD control utilizing an equidistant time grid with $\Delta t = 1/40$ (middle) and $\ell = 4$, POD control utilizing an adaptive time grid with dof = 41 (right) and $\ell = 4$.

TABLE 13. Test 2: absolute errors between the analytical optimal solution and the POD solution with $\ell = 1$ depending on the time discretization (equidistant: columns 1–4, adaptive: columns 5–8).

Δt	$\varepsilon^y_{ m abs}$	$\varepsilon^u_{ m abs}$	$\varepsilon^p_{ m abs}$	dof	$\varepsilon^y_{ m abs}$	$\varepsilon^u_{ m abs}$	$\varepsilon^p_{ m abs}$
1/20	$5.0631 \cdot 10^{-01}$	$7.8420 \cdot 10^{+00}$	$3.5413 \cdot 10^{+01}$	21	$4.5255 \cdot 10^{-02}$	$5.4054 \cdot 10^{-01}$	$2.4409 \cdot 10^{+00}$
1/40	$2.6230 \cdot 10^{-01}$	$4.1059 \cdot 10^{+00}$	$1.8542 \cdot 10^{+01}$	41	$2.0721 \cdot 10^{-02}$	$5.3475 \cdot 10^{-03}$	$1.3186 \cdot 10^{-02}$
1/68	$1.5684 \cdot 10^{-01}$	$2.4503 \cdot 10^{+00}$	$1.1065 \cdot 10^{+01}$	69	$2.0713 \cdot 10^{-02}$	$4.5706 \cdot 10^{-03}$	$4.2670 \cdot 10^{-03}$
1/134	$8.1129 \cdot 10^{-02}$	$1.2386 \cdot 10^{+00}$	$5.5938 \cdot 10^{+00}$	135	$2.0664 \cdot 10^{-02}$	$4.4905 \cdot 10^{-03}$	$2.3507 \cdot 10^{-03}$

TABLE 14. Test 3: inclusion of box constraints for the control intensities: absolute errors between the analytical optimal solution and the POD solution with $\ell = 4$ depending on the time discretization (equidistant: columns 1–4, adaptive: columns 5–8).

Δt	$\varepsilon^y_{ m abs}$	$\varepsilon^u_{ m abs}$	$\varepsilon^p_{ m abs}$	dof	$\varepsilon^y_{ m abs}$	$\varepsilon^u_{ m abs}$	$\varepsilon^p_{ m abs}$
1/20	$2.8601 \cdot 10^{-01}$	$5.7201 \cdot 10^{+00}$	$3.5430 \cdot 10^{+01}$	21	$2.2714 \cdot 10^{-02}$	$3.9586 \cdot 10^{-01}$	$2.4423 \cdot 10^{+00}$
1/40	$1.4802 \cdot 10^{-01}$	$2.9955 \cdot 10^{+00}$	$1.8551 \cdot 10^{+01}$	41	$2.9482 \cdot 10^{-04}$	$4.4969 \cdot 10^{-03}$	$1.3183 \cdot 10^{-02}$
1/68	$8.8124 \cdot 10^{-02}$	$1.7882 \cdot 10^{+00}$	$1.1071 \cdot 10^{+01}$	69	$2.1247 \cdot 10^{-04}$	$3.2811 \cdot 10^{-03}$	$4.2629 \cdot 10^{-03}$
1/134	$4.4570 \cdot 10^{-02}$	$9.0470 \cdot 10^{-01}$	$5.5965 \cdot 10^{+00}$	135	$2.1330 \cdot 10^{-04}$	$3.1321 \cdot 10^{-03}$	$2.3474 \cdot 10^{-03}$
1/6500	$2.6019 \cdot 10^{-04}$	$3.8720 \cdot 10^{-03}$	$9.3290 \cdot 10^{-03}$	_	_	_	_



FIGURE 13. Test 3: inclusion of box constraints for the control intensities: analytical control intensities $\bar{u}_1(t)$ (top left) and $\bar{u}_2(t)$ (bottom left), POD control utilizing an equidistant time grid with $\Delta t = 1/40$ (middle) and $\ell = 4$, POD control utilizing an adaptive time grid with dof = 41 (right) and $\ell = 4$.

use of more than $\ell = 4$ POD basis functions does not lead to more accurate approximations. The first POD mode ψ_1 can be seen in the right panel of Figure 8. Table 13 shows the absolute error between the analytical solution and the POD solution in the state, control and adjoint state for uniform as well as for adaptive time discretization for $\ell = 1$. We note that in the case of the uniform temporal discretization, the use of $\ell = 1$ POD basis function leads to similar approximation results as using $\ell = 4$ POD modes. On the contrary, the adaptive time discretization with $\ell = 4$ POD basis functions leads to more accurate approximation results for the state variable than using $\ell = 1$ POD modes (compare Tabs. 13 and 10). However, with only one POD mode, the time adaptive grid gives very accurate results.

6.3. Test 3: control constrained problem

In this test we add control constraints to the previous example. We set $u_{1,a}(t) \leq u_1(t) \leq u_{1,b}(t)$ and $u_{2,a}(t) \leq u_2(t) \leq u_{2,b}(t)$ for the time dependent control intensities $u_1(t)$ and $u_2(t)$. The analytical value range for both controls is $u_1(t), u_2(t) \in [-0.3479, 0.1700]$ for $t \in [0, 1]$. For each control intensity we choose different upper and lower bounds: we set $u_{1,a}(t) = -100$ (*i.e.* no restriction), $u_{1,b} = 0.1$ and $u_{2,a}(t) = -0.2$, $u_{2,b}(t) = 0$. For the solution of problem (4.5) we use a projected gradient method. For both the full solution and the reduced order solution, 5 projected gradient steps are needed.

The solution of the nonlinear, nonsmooth equation (3.1) can be done by a semi-smooth Newton method or by a Newton method utilizing a regularization of the projection formula, see [24]. In our numerical tests we compute the approximate solution to (3.1) with a fixed point iteration and initialize the method with the adjoint state corresponding to the control unconstrained optimal control problem. In this way, only two iterations are needed for convergence. Convergence of the fixed point iteration can be argued for large enough values of α , see [12].

The analytical optimal solutions \bar{u}_1 and \bar{u}_2 are shown in the left plots in Figure 13. For POD basis computation, we use state, adjoint and time derivative adjoint snapshots corresponding to the reference control $u_0 = 0$

Δt	POD offline	POD online	dof	Compute \mathcal{T}	POD offline	POD online
1/20	$0.16\mathrm{s}$	$0.06\mathrm{s}$	21	$1.09\mathrm{s}$	$0.16\mathrm{s}$	0.06 s
1/40	$0.17\mathrm{s}$	$0.12\mathrm{s}$	41	$5.31\mathrm{s}$	$0.17\mathrm{s}$	$0.12\mathrm{s}$
1/68	$0.20\mathrm{s}$	$0.14\mathrm{s}$	69	$7.84\mathrm{s}$	$0.20\mathrm{ss}$	$0.14\mathrm{s}$
1/134	$0.25\mathrm{s}$	$0.22\mathrm{s}$	135	$13.09\mathrm{s}$	$0.25\mathrm{s}$	$0.22\mathrm{s}$
1/6500	$46.76\mathrm{s}$	$9.12\mathrm{s}$	—	_	_	_

TABLE 15. Test 3: CPU times for POD offline computations (snapshot generation and POD basis computation) depending on the time discretization (equidistant: columns 1–3, adaptive: columns 4–7).

and we also include the initial condition y_0 into our snapshot set. Figure 13 refers to the POD controls using a uniform (*middle panel*) and an adaptive temporal discretization (*right panel*). We again like to emphasize how accurate our approximation with an adaptive time grid is in comparison to a uniform grid (see Tab. 14). We note that the inclusion of box constraints on the control functions leads in general to better approximation results, compare Table 10 with Table 14. This is due to the fact that on the active sets the error between the analytical optimal controls and the POD solutions vanishes.

The CPU time is listed in Table 15. As one can see, to achieve an accuracy of order 10^{-3} for the control and adjoint variable, $n \approx 6500$ time steps are necessary with a uniform temporal discretization (compare Tab. 14). In this case, the POD simulation including the offline phase takes 55.88 s, whereas utilizing our approach with 69 degrees of freedom takes around 8.18 s. This gives us an impressive speedup of factor approximately 7.

7. CONCLUSION

In this paper we investigated the problem of snapshot location in optimal control problems. We showed that the numerical POD solution is much more accurate if we use an adaptive time grid, especially when the solution of the problem presents steep gradients. The time grid was computed by means of an *a posteriori* error estimation strategy of a space-time approximation of a second order in time and fourth order in space elliptic equation which describes the optimal control problem and has the advantage that it is independent of an input control function. Furthermore, a coarse approximation with respect to space of the latter equation gives information about the snapshots one can use to build the surrogate model. Finally, we provided a certification of our surrogate model by means of an *a posteriori* error estimation for the error between the optimal solution and the POD solution.

In the next step, we extend our approach to optimal control problems subject to nonlinear parabolic equations. In case of fully distributed controls, a reformulation of the optimality system as a second order in time and fourth order in space nonlinear elliptic PDE for the state is also possible. However, the reformulation with respect to p still would contain the state y, so that our approach does not directly extend to this situation. But, if one applies the SQP framework to the solution of the underlying optimal control problem our approach is directly applicable to the optimality conditions associated with the linear quadratic SQP subproblems.

References

- K. Afanasiev and M. Hinze, Adaptive Control of a Wake Flow Using Proper Orthogonal Decomposition. Vol. 216 of Lecture Notes in Pure and Applied Mathematics. Shape Optimization & Optimal Design, Marcel Dekker (2001) 317–332.
- [2] A. Alla and M. Falcone, An Adaptive POD Approximation Method for the Control of Advection-Diffusion Equations. Vol. 164 of Control and Optimization with PDE Constraints, edited by K. Kunisch, K. Breides, C. Clason, G. von Winckel. International Series of Numerical Mathematics, Birkhäuser, Basel (2013) 1–17.
- [3] A. Alla, C. Gräßle and M. Hinze, A Residual Based Snapshot Location Strategy for POD in Distributed Optimal Control of Linear Parabolic Equations. Vol. 49 of 2nd IFAC Workshop on Control of Systems Governed by Partial Differential Equations CPDE 2016 (2016) 13–18.
- [4] E. Arian, M. Fahl and E. Sachs, Trust-region proper orthogonal decomposition models by optimization methods. In Proc. of the 41st IEEE Conference on Decision and Control, Las Vegas, Nevada (2002) 3300–3305.

- [5] L.C. Evans, Partial Differential Equations. Vol. 19 of Graduate Studies in Mathematics. American Mathematical Society, Providence, RI (2010).
- [6] J. Ghiglieri and S. Ulbrich, Optimal flow control based on POD and MPC and an application to the cancellation of tollmienschlichting waves. Optim. Methods Softw. 29 (2014) 1042–1074.
- [7] W. Gong, M. Hinze and Z.J.Zhou, Space-time finite element approximation of parabolic optimal control problems. J. Numer. Math. 20 (2012) 111–145.
- [8] C. Gräßle, M. Gubisch, S. Maier, S. Rogg and S. Volkwein, POD basis updates for nonlinear PDE control. Automatisierungstechnik 65 (2017) 298–307.
- [9] M. Gubisch and S. Volkwein, Proper Orthogonal Decomposition for Linear-Quadratic Optimal Control. Model Reduction and Approximation: Theory and Algorithms, edited by P. Benner, A. Cohen, M. Ohlberger and K. Willcox. SIAM, Philadelphia, PA (2017).
- [10] M. Gubisch, I. Neitzel and S. Volkwein, Model reduction and approximation: theory and algorithms. Computational Science and Engineering. SIAM (2016).
- M. Hinze, A variational discretization concept in control constrained optimization: the linear-quadratic case. Comput. Optim. Appl. 30 (2005) 45–61.
- [12] M. Hinze and M. Vierling, Variational discretization and semi-smooth Newton methods; implementation, convergence and globalization in PDE constrained optimization with control constraints. Optim. Methods Softw. 27 (2012) 933–950.
- [13] M. Hinze and S. Volkwein, Error estimates for abstract linear-quadratic optimal control problems using proper orthogonal decomposition. S. Comput. Optim. Appl. 39 (2008) 319–345.
- [14] M. Hinze, R. Pinnau, M. Ulbrich and S. Ulbrich, Optimization with PDE Constraints. Vol. 23 of Mathematical Modelling: Theory and Applications. Springer Verlag (2009).
- [15] R.H.W. Hoppe and Z. Liu, Snapshot location by error equilibration in proper orthogonal decomposition for linear and semilinear parabolic partial differential equations. J. Numer. Math. 22 (2014) 1–32.
- [16] E. Kammann, F. Tröltzsch and S. Volkwein, A method of a posteriori error estimation with application to proper orthogonal decomposition. ESAIM: M2AN 47 (2013) 555–581.
- [17] Z. Kanar Seymen, H. Yücel and B. Karasözen, Distributed optimal control of time-dependent diffusion-convection-reaction equations using space-time discretization. J. Comput. Appl. Math. 261 (2014) 146–157.
- [18] K. Kunisch and S. Volkwein, Galerkin proper orthogonal decomposition methods for parabolic problems. Numer. Math. 90 (2001) 117–148.
- [19] K. Kunisch and S. Volkwein, Galerkin proper orthogonal decomposition methods for a general equation in fluid dynamics. SIAM, J. Numer. Anal. 40 (2002) 492–515.
- [20] K. Kunisch and S. Volkwein, Proper orthogonal decomposition for optimality systems. ESAIM: M2AN 42 (2008) 1–23.
- [21] K. Kunisch and S. Volkwein, Optimal Snapshot Location for computing POD basis functions. ESAIM: M2AN 44 (2010) 509–529.
- [22] J.L. Lions, Optimal Control of Systems Governed by Partial Differential Equations. *Grundlehren der mathematischen Wissenschaften*. Springer (1971).
- [23] I. Neitzel and B. Vexler, A priori error estimates for space-time finite element discretization of semilinear parabolic optimal control problems. Numerische Mathematik 120 (2012) 345–386.
- [24] I. Neitzel, U. Prüfert and T. Slawig, A smooth regularization of the projection formula for constrained parabolic optimal control problems. Numer. Funct. Anal. Optim. 32 (2011) 1283–1315.
- [25] J. Nocedal and S.J. Wright, Numerical Optimization, second edition. Springer Series in Operation Research (2006).
- [26] N.C. Nguyen, G. Rozza and A.T. Patera, Reduced basis approximation and a posteriori error estimation for time dependent viscous Burgers equation. Calcolo 46 (2009) 157–185.
- [27] G.M. Oxberry, T. Kostova-Vassilevska, B. Arrighi and K. Chand, Limited-memory adaptive snapshot selection for proper orthogonal decomposition. Int. J. Numer. Methods Eng. 109 (2017) 198–217.
- [28] A.T. Patera and G. Rozza, Reduced Basis Approximation and A posteriori Error Estimation for Parametrized Partial Differential Equations. *MIT Pappalardo Graduate Monographs in Mechanical Engineering*. Oxford University (2006).
- [29] G. Rozza, D.B.P. Huynh and A.T. Patera, Reduced basis approximation and a posteriori error estimation for affinely parametrized elliptic coercive partial differential equations. Arch. Comput. Methods. Eng. 15 (2008) 229–275.
- [30] L. Sirovich, Turbulence and the dynamics of coherent structures. Parts I–II. Q. Appl. Math. XVL (1987) 561–590.
- [31] F. Tröltzsch, Optimal control of partial differential equations: theory, methods and application. Am. Math. Soc. (2010).
- [32] F. Tröltzsch and S. Volkwein, POD a posteriori error estimates for linear-quadratic optimal control problems. Comput. Optim. Appl. 44 (2009) 83–115.
- [33] S. Volkwein, Optimality system POD and a posteriori error analysis for linear-quadratic problems. Control Cybern. 40 (2011) 1109–1125.
- [34] S. Volkwein and A. Studinger, Numerical Analysis of POD a posteriori error estimation for optimal control. Control and Optimization with PDE Constraints, edited by K. Kunisch, K. Breides, C. Clason and G. von Winckel. In Vol. 164 of *International Series of Numerical Mathematics*. Birkhäuser, Basel (2013) 137–158.