Institut für Numerische Simulation
Rheinische Friedrich-Wilhelms-Universität Bonn

Wegelerstraße 6 • 53115 Bonn • Germany
phone +49 228 73-3427 • fax +49 228 73-7527
www.ins.uni-bonn.de

M. Gubisch, I. Neitzel, S. Volkwein

A-posteriori error estimation of discrete POD models for PDE-constrained optimal control

INS Preprint No. 1608

March 2016

A-posteriori error estimation of discrete POD models for PDE-constrained optimal control

Martin Gubisch, Ira Neitzel, and Stefan Volkwein

Abstract In this work a-posteriori error estimates for linear-quadratic optimal control problems governed by parabolic equations are considered. Different error estimation techniques for finite element discretizations and model-order reduction are combined to validate suboptimal control solutions from low-order models which are constructed by Galerkin discretization and application of proper orthogonal decomposition (POD). The theoretical findings are used to design an efficient updating algorithm for the reduced-order models; the efficiency and accuracy is illustrated by numerical experiments.

1 Introduction

Many optimal control problems with partial differential equations (PDEs), especially those in higher dimensions, are challenging to be solved numerically because their discretization leads to very high-dimensional problems. This is the reason why model reduction techniques, such as the method of proper orthogonal decomposition (POD), are subject to active research.

The method of POD approximates a high-dimensional problem by a smaller, tractable problem by means of projections of the dynamical system onto subspaces that inherit characteristics of the expected solution. Regarding convergence results

Martin Gubisch
Universität Konstanz, Fachbereich Mathematik und Statistik, Universitätsstraße 10, 78457 Konstanz, Germany, e-mail: martin.gubisch@uni-konstanz.de

Ira Neitzel
Rheinische Friedrich-Wilhelms-Universität Bonn, Institut für Numerische Simulation, Wegelerstr. 6, 53115 Bonn, Germany, e-mail: neitzel@ins.uni-bonn.de

Stefan Volkwein
Universität Konstanz, Fachbereich Mathematik und Statistik, Universitätsstraße 10, 78457 Konstanz, Germany, e-mail: stefan.volkwein@uni-konstanz.de

for POD solutions to parabolic PDEs we refer, for instance, to [15, 16]. In [10], an overview over the topic of POD model order reduction is provided.

Recently, some effort has been made to derive a-priori and a-posteriori error analysis for the reduced control problems. We refer to [12] for a-priori error estimates for POD approximations to control problems, and to [29] for first results on a-posteriori error estimates for linear-quadratic problems. These results are extended to non-convex problems in [14], and to problems with mixed control-state constraints in [9]. The numerical analysis of POD a-posteriori error estimation for optimal control problems is investigated in [28].

However, the available results in the literature on POD a-posteriori error estimates so far do not account for the fact that the subspace for the reduced model is generated from snapshots of the full order model, which, in the setting of PDE constrained optimization, is typically a finite element discretization of a continuous model problem. Thus, what is commonly referred to as the *true solution* in the context of reduced order models (ROM) is itself in fact an approximation of the *real solution*.

For the finite element approximation of the real solution of parabolic problems, there exists a range of results on a-priori and a-posteriori error estimates. Concerning a-priori estimates, we refer to [23] and references therein, where error estimates are provided for a space-time finite element discretization of PDE-constrained optimal control problems without further control or state constraints. This approach has been extended to problems with additional control constraints in [24], that – after minor modifications to the simplified structure of the control space – covers the model problem to be discussed in our paper. The discretization therein includes a (discontinuous Galerkin type) variant of the implicit Euler scheme for the time discretization and usual H^1-conforming finite element discretization in space. We will heavily rely on this Galerkin structure of the discretization in Sec. 4, where we derive our main result. We note in passing that finite element discretization error estimates for parabolic problems are available also for certain types of state constraints, see [25] or [8], or semilinear parabolic problems with pointwise-in-time state constraints, cf. [20], or control constraints, see e.g. [27]. Cf. also results on plain convergence without any rates in [2].

However, standard a-priori estimates are in general not a good indicator for enlarging or updating the reduced-order models since they usually tend to overestimate the effective errors significantly. Instead, a-posteriori error estimates may be applied; [1] provides such methods for uncontrolled parabolic PDEs, focussing on adaptive time and spatial grid selection. In [13], a detailled discussion about the discretization and regularization of optimal control problems is presented. For adaptive discretization strategies based on a-posteriori error estimates for parabolic optimal control without additional inequality constraints we refer to e.g. [22] and [26] and the references therein. In [17], the authors introduce a POD basis updating strategy which includes the POD eigenvalue problem (and hence also the high-order state equation) into the ROM framework and applies a variable splitting technique to devide the resulting mixed optimality system into its high- and low-dimensional part. A strategy to locate optimal snapshots of some PDE solution can be found in [18].

In this paper, we are concerned with the following model problem with state y and control u in spaces to be specified,

$$\text{Minimize } J(y,u) := \frac{1}{2}\int_0^T \int_\Omega (y(t,x)-y_d(t,x))^2 dx\,dt + \frac{\nu}{2}\sum_{i=1}^{N_u}\int_0^T u_i(t)^2\,\mathrm{d}t \tag{1a}$$

subject to

$$\partial_t y - \Delta y = \sum_{i=1}^{N_u} u_i\chi_i + f \text{ in } (0,T)\times\Omega, \qquad y(0,\cdot) = y_0 \text{ in } \Omega, \tag{1b}$$

$$y = 0 \text{ in } (0,T)\times\partial\Omega, \qquad u_a \le u(t) \le u_b \text{ almost everywhere in } (0,T) \tag{1c}$$

where the last inequality is to be understood componentwise. Note that the functions $\chi_i : \Omega \to \mathbb{R}$, $i = 1,\dots,N_u$, are given, fixed data. The exact setting of the model problem will be described in the next section.

We are interested in discussing how the discretization error on the one hand and the model order reduction error on the other hand relate and can possibly be balanced, motivated by the following two questions:

1. Since a reduced-order model is based on the snapshots of a high-dimensional approximative PDE solution, the model already includes the finite element discretization errors and it does not make sense to decrease the POD residual below the order of the finite element discretization error.
2. If, however, the error of the POD approximation does not reflect the order of the discretization error even when increasing the POD basis rank, i.e. the size of the reduced order model, the current POD basis may not reflect the dynamics of the optimal POD basis $\bar{\psi}$ refering to the optimal state solution $\bar{y}$. In this case, an update of the POD basis may be required to improve the results.

The paper is organized as follows: In Sec. 2, we summarize available results for the continuous optimal control problem, such as existence and regularity results as well as optimality conditions. Then, in Sec. 3, we briefly describe the finite element discretization along the lines of [24], and state an a-priori error estimate for the finite element discretization from [24]. In short, the discretization is utilized by discretizing the state and adjoint equations by the so-called dG(0)cG(1) method, cf. for instance [5, 6]. That is, the time-discretization of the PDEs is done by piecewise constant functions, whereas usual H^1-conforming finite elements in space are used. The time-dependent controls are discretized implicitly via the optimality conditions, cf. also the variational discretization approach in [11]. For convenience of the reader, we summarize existence and regularity results for the solution of the semidiscrete and fully discrete state equations, as well as existence of unique solutions to the optimal control problems and optimality conditions on each discretization level. For our main result in the subsequent section, we will use in particular the optimality conditions from Sec. 3.2 as well as the error estimate from [24], which we state in Sec. 3.3. Our main result, the a-posteriori error analysis for a suboptimal discrete

solution, where we have in mind the discrete POD solution, follows in Sec. 4. The main step in our analysis is to extend the a-posteriori error analysis technique from [29], which is related to the one in [21] for ordinary differential equations, to estimate the error between the full-order finite element solution and the (discrete) solution of the reduced model. Here, we readily make use of the fact that the dG(0)cG(1) discretization is a Galerkin scheme. We eventually obtain an a-posteriori error estimator along the lines of [29], but in contrast to the latter paper, the estimator is computable in the strict sense since it only depends on the computed, finite element solution rather than the true unknown continuous solution. Comparing this error to the finite element error can be an indication if for instance a POD basis update is usefull. Then, Sec. 5 finally describes the method of model order reduction via proper orthogonal decomposition. We end this paper by numerical experiments in Sec. 6, where we also address some aspects of implementation, such as estimating the constants in the a-priori error estimates.

Let us remark that for future analysis, it seems also reasonable to include the discussion of a-posteriori finite element error analysis. However, the focus of our paper is on a computable a-posteriori error estimate for the solution of the reduced-order model. Since we clearly separate the influence of discretization and model-order reduction errors, these results may readily be extended to balance this error with any type of available a-posteriori discretization error estimate.

2 Optimization problem

In the following, we lay out the principle assumptions on the data in (1) and summarize known results on existence and regularity of solutions to the underlying PDE, the control problem itself, as well as first-order necessary and, due to convexity, also sufficient optimality conditions.

Assumption 2.1 *Let $\Omega \subset \mathbb{R}^n$, $n = 1,2,3$, be a convex polygonal or polyhedral domain with boundary $\partial\Omega$ for $n = 2,3$, or an open interval in $\mathbb{R}$. Moreover, let $T > 0$ be a given real number that defines the time interval $I := (0,T)$. In addition, $\nu \in \mathbb{R}$ is a positive, fixed parameter, and the bounds $u_a, u_b \in \mathbb{R}^{N_u}$ are vectors of real numbers that fulfill $u_a < u_b$ componentwise. The desired state y_d is a function from $L^2(I \times \Omega)$ and the initial state y_0 is a function from $H_0^1(\Omega)$. For the shape functions $\chi_i \colon \Omega \to \mathbb{R}$, $i = 1,\dots,N_u$, we require $\chi_i \in H_0^1(\Omega)$.*

We introduce the following short notation for inner products and norms on the spaces $L^2(\Omega)$ and $L^2(I \times \Omega)$, as well as $L^2(I;\mathbb{R}^{N_u})$:

$$(v,w) := (v,w)_{L^2(\Omega)}, \qquad (v,w)_I := (v,w)_{L^2(I\times\Omega)}, \qquad \langle v,w\rangle_I := \langle v,w\rangle_{L^2(I;\mathbb{R}^{N_u})}$$
$$\|v\| := \|v\|_{L^2(\Omega)}, \qquad \|v\|_I := \|v\|_{L^2(I\times\Omega)}, \qquad |v|_I := \|v\|_{L^2(I;\mathbb{R}^{N_u})}.$$

Throughout the paper we abbreviate $V := H_0^1(\Omega)$; $c > 0$ will denote generic auxiliary constants. Moreover, in order to find a weak formulation of the state equation

(1b) and the optimal control problem (1), we introduce the state space Y, the control space U, and the set of admissible controls U_{ad},

$$Y := W(0,T) = \{v \mid v \in L^2(I,V) \text{ and } \partial_t v \in L^2(I,V^*)\}, \qquad U := L^2(I,\mathbb{R}^{N_u}),$$
$$U_{\text{ad}} := \{u \in U \mid u_a \le u(t) \le u_b \text{ for a.a. } t \in I \text{ componentwise}\},$$

as well as the control operator

$$B\colon U \to L^2(I\times\Omega), \quad u \mapsto \sum_{i=1}^{N_u} u_i(t)\chi_i(x).$$

A weak formulation of the state equation (1b) for a fixed control $u \in U$ and fixed initial state $y_0 \in V$ as well as source term $f \in L^2(I\times\Omega)$ is to find a state $y \in Y$ that satisfies

$$\int_0^T (\partial_t y, \varphi)_{V^*,V}\,\mathrm{d}t + (\nabla y, \nabla\varphi)_I = (Bu,\varphi)_I + (f,\varphi)_I \;\forall \varphi \in L^2(I,V),\; y(0,\cdot) = y_0. \quad (2)$$

The following existence and regularity result is readily available from [7].

Proposition 1. *For fixed control $u \in U$, fixed source term $f \in L^2(I\times\Omega)$, and fixed initial state $y_0 \in V$ there exists a unique solution $y \in Y$ of the weak state equation* (2)*. Moreover, the solution exhibits the improved regularity*

$$y \in L^2(I,H^2(\Omega)\cap V)\cap H^1(I,L^2(\Omega)) \hookrightarrow C(\bar{I},V)$$

and the stability estimate

$$\|\partial_t y\|_I + \|y\|_I + \|\nabla y\|_I + \|\nabla^2 y\|_I \le C(|u|_I + \|f\|_I + \|\nabla y_0\|) \quad (3)$$

is satisfied for a constant $C > 0$.

By the regularity $y \in H^1(0,T;L^2(\Omega))$ it is justified to use the bilinear form

$$\mathbf{b}(y,\varphi) := (\partial_t y,\varphi)_I + (\nabla y,\nabla\varphi)_I,$$

and use the weak formulation

$$\mathbf{b}(y,\varphi) = (Bu,\varphi)_I + (f,\varphi)_I \quad \forall \varphi \in Y, \qquad y(0,\cdot) = y_0. \quad (4)$$

Note that due to the linearity of the state equation, (1) can be reformulated equivalently into a setting with homogeneous initial condition and without additional source term f by splitting the solution of (4) into two parts $y = \hat{y} + y_u$, which fulfill the PDEs

$$\partial_t \hat{y} - \Delta\hat{y} = f \text{ in } I\times\Omega, \quad \hat{y}(0,\cdot) = y_0 \text{ in } \Omega, \quad \hat{y} = 0 \text{ in } I\times\partial\Omega, \quad (5)$$

as well as

$$\partial_t y_u - \Delta y_u = \sum_{i=1}^{N_u} u_i(t)\chi_i(x) \text{ in } I\times\Omega, \quad y_u(0,\cdot) = 0 \text{ in } \Omega, \quad y_u = 0 \text{ in } I\times\partial\Omega \quad (6)$$

in the weak sense. The fixed term $\hat{y}$ is independent of the controls and can be incorporated into the desired state y_d. For ease of presentation, we will therefore assume without loss of generality: $y_0 = 0$ as well as $f = 0$.

In the following we will, however, state more general results from [23, 24] for nonhomogeneous initial conditions and additional source terms.

Next we introduce the linear control-to-state mapping $S : U \to Y$, $Su = y_u$, which leads to the reduced objective function $\hat{J} : U \to \mathbb{R}_0^+$ with $u \mapsto J(S(u),u)$. Note that here and in the following, we tacitly use the operator S also if we interpret the state y as a function in $L^2(I\times\Omega)$. This makes the optimal control problem (1) equivalent to

$$\text{Minimize } \hat{J}(u) \text{ subject to } u \in U_{\text{ad}}. \quad (\mathbf{P})$$

The following existence and uniqueness result is obtained by standard arguments, cf. for instance [30], since the set of admissible controls is not empty by Ass. 2.1.

Lemma 1. *Let Assumption 2.1 be satisfied. Then the optimal control problem* (**P**) *admits a unique optimal control $\bar{u} \in U_{\text{ad}}$ with associated optimal state $\bar{y} = S\bar{u}$.*

Let us refer to [30] for a detailed proof. We proceed by discussing standard first order necessary optimality conditions for the optimal control problem with the help of a variational inequality. Due to convexity, these conditions are also sufficient for optimality.

Lemma 2. *A control $\bar{u}$ is the unique solution of* (**P**) *if and only if $\bar{u} \in U_{\text{ad}}$ and the following variational inequality holds:*

$$\hat{J}'(\bar{u})(u-\bar{u}) \geq 0 \quad \forall u \in U_{\text{ad}}. \quad (7)$$

For a proof, we refer again to e. g., [30]. In order to express the optimality conditions in a more convenient way we define for any control $u \in U_{\text{ad}}$ the adjoint state variable $p = p(u) \in Y$, which is the solution of

$$-\partial_t p - \Delta p = y_d - y(u) \text{ in } I\times\Omega, \quad p(T,\cdot) = 0 \text{ in } \Omega, \quad p = 0 \text{ in } I\times\partial\Omega \quad (8)$$

with $y(u) = Su$. A weak solution of this adjoint problem can be defined by means of the already introduced bilinear form $\mathbf{b}$, since elements of Y can be integrated by parts in time. We obtain

$$\mathbf{b}(\varphi,p) = (\varphi, y_d - y(u))_I \quad \forall \varphi \in Y, \qquad p(T,\cdot) = 0. \quad (9)$$

Note that Prop. 1 is applicable to (9) after a time transformation $\tau = T - t$. We then rewrite the first-order optimality conditions from Prop. 2 in the form

$$\langle \nu\bar{u} - B^* p(\bar{u}), u-\bar{u}\rangle_I \geq 0 \qquad \forall u \in U_{\text{ad}},$$

where $B^*: L^2(I\times\Omega)\to U$ denotes the Hilbert space adjoint operator of B satisfying the formula $(Bu,v)_I = \langle u, B^*v\rangle_I$ for all $u,v\in U$. The following identity for B^* can easily be verified by means of the above definition. For any $\varphi\in L^2(I\times\Omega)$, we find

$$B^*\varphi = u, \qquad u_i(t) := \int_\Omega \varphi(t,x)\chi_i(x)dx, \quad i=1,\dots,N_u \text{ and } t\in I.$$

Then, using the pointwise projection on the admissible set,

$$P_{\text{ad}}: U\to U_{\text{ad}}, \quad P_{\text{ad}}(u)_i(t) := \max(u_a,\min(u_b,u_i(t))), \quad i=1,\dots,N_u,$$

the optimality condition simplifies further to

$$\bar u = P_{\text{ad}}\left(\frac{1}{\nu}B^*p(\bar u)\right). \tag{10}$$

We refer to [30] for the technique of proof. From Prop. 1 and the projection formula, we deduce the following regularity result, which is a direct consequence of Prop. 2.3 in [24] taking into account that the controls depend only on the time variable.

Proposition 2. *Let $\bar u$ be the solution of the optimization problem* (**P**) *with associated state $\bar y = y(\bar u)$ and let $\bar p := p(\bar u)$ denote the corresponding adjoint state. Then $\bar u,\bar y,\bar p$ achieve the following regularities:*

$$\bar y,\bar p\in L^2(I,H^2(\Omega)\cap V)\cap H^1(I,L^2(\Omega))\hookrightarrow C(\bar I,V),\ \bar u\in H^1(I;\mathbb{R}^{N_u})\hookrightarrow C(\bar I;\mathbb{R}^{N_u})$$

3 Discretization of the problem

This section is devoted to the finite element discretization of the optimal control problem under consideration. We review the discretization procedure as well as results on e.g. stability of discrete solutions, and an a-priori error estimate for the controls on the continuous and discrete level from e.g. [24] for linear-quadratic problems with controls varying in space and time, or [27], where the state equation is nonlinear, but the setting of only time-dependent controls is addressed explicitely. More precisely, we first give a brief overview about semidiscretization of the state (and adjoint) equation in time by piecewise constant functions, with values in V. Note that the resulting scheme is a variant of the implicit Euler method. Even though the control functions themselves will not be discretized explicitely, we will obtain, by means of the semidiscrete optimality conditions, that the optimal control is in fact piecewise constant in time, and therefore a discrete function. In a second step, we discretize the involved PDEs also in space. Here, we use usual H^1-conforming linear finite elements. The resulting discretization scheme is commonly referred to as dG(0)cG(1) method. Since the control functions are functions in time, only, solving the time-and-space discrete optimality system corresponds to solving a completely discretized problem.

3.1 Semidiscretization in time

Along the lines of [23, 24], let a partitioning of the time interval $\bar{I} = [0,T]$ be given as $\bar{I} = \{0\} \cup I_1 \cup I_2 \cup \ldots \cup I_M$ with subintervals $I_m = (t_{m-1}, t_m]$ of size k_m, defined by time points $0 = t_0 < t_1 < \ldots < t_{M-1} < t_M = T$. The discretization parameter k is defined as a piecewise constant function by setting $k_{|I_m} = k_m$ for $m = 1,2,\ldots,M$, yet k also denotes the maximal size of the time steps, i.e., $k = \max k_m$. The semidiscrete trial and test space is given by

$$Y_k = \{v_k \in L^2(I,V) \mid v_{k|I_m} \in \mathscr{P}_0(I_m,V),\ m = 1,2,\ldots,M\},$$

where $\mathscr{P}_0(I_m,V)$ denotes the space of constant functions defined on I_m with values in V. The control space U and the set of admissible controls U_{ad} remain unchanged, yet we will later find that the semidiscrete optimal control is an element of the space

$$U_k := \{v_k \in U \mid v_{k|I_m} \in \mathscr{P}_0(I_m,\mathbb{R}^{N_u}),\ m = 1,2,\ldots,M\}.$$

For functions $v_k \in Y_k$ we define

$$\begin{aligned} v_{k,m}^+ &:= \lim_{t\to 0^+} v_k(t_m + t) = v_k(t_{m+1}) =: v_{k,m+1}, \\ v_{k,m}^- &:= \lim_{t\to 0^+} v_k(t_m - t) = v_k(t_m) =: v_{k,m}, \\ [v_k]_m &:= v_{k,m}^+ - v_{k,m}^- = v_{k,m+1} - v_{k,m} \end{aligned}$$

and introduce the short notation

$$(v,w)_{I_m} := (v,w)_{L^2(I_m\times\Omega)}, \qquad \|v\|_{I_m} := \|v\|_{L^2(I_m\times\Omega)}$$

for functions $v,w \in L^2(I_m \times \Omega)$. The semidiscrete version of the bilinear form $\mathbf{b}(\cdot,\cdot)$ for $y_k, \varphi_k \in Y_k$ is given by

$$\mathbf{b}_k(y_k,\varphi_k) = (\nabla y_k, \nabla\varphi_k)_I + \sum_{m=2}^{M} (y_{k,m} - y_{k,m-1}, \varphi_m) + (y_{k,1}, \varphi_{k,1}),$$

and the dG(0) semidiscretization of the state equation (4) for fixed control $u \in U$ reads as follows: Find a state $y_k = y_k(u) \in Y_k$ such that

$$\mathbf{b}_k(y_k,\varphi_k) = (Bu,\varphi_k)_I + (f,\varphi_k)_I + (y_0,\varphi_{k,1}) \quad \forall \varphi_k \in Y_k. \tag{11}$$

Here, $\varphi_{k,m}$ denotes $\varphi_{k_{|I_m}}$. Note that we assumed $y_0 = 0$ and $f = 0$ for ease of presentation, but include the more general setting in the following theorem.

Theorem 1. *For every fixed control $u \in U$, the semidiscrete state equation* (11) *with potentially nonhomogenous initial state $y_0 \in V$ and source term $f \in L^2(I \times \Omega)$ admits a unique semidiscrete solution $y_k \in Y_k$ satisfying the stability result*

$$\|y_k\|_I^2 + \|\nabla y_k\|_I^2 + \|\Delta y_k\|_I^2 + \sum_{i=1}^{M} k_m^{-1} \|y_{k,i} - y_{k,i-1}\|^2 \leq C\{|u|_I^2 + \|f\|_I^2 + \|\nabla y_0\|^2\}$$

with a constant $C > 0$ independent of the discretization parameters.

Proof. This is a direct consequence of Theorem 4.1 in [23], taking into account that the controls are only time-dependent functions.

With the semidiscrete control-to-state operator $S_k : U \to Y_k$, $S_k(u) = y_k$, where y_k is the solution of (11), and consequently a semidiscrete reduced objective function

$$\hat{J}_k : U \to \mathbb{R}_0^+, \quad u \mapsto J(S_k(u), u),$$

the reduced semidiscrete problem formulation reads

$$\text{Minimize } \hat{J}_k(u) \text{ subject to } u \in U_{\text{ad}}. \tag{$\mathbf{P}_k$}$$

Existence of a unique semidiscrete optimal control $\bar{u}_k \in U_{\text{ad}}$ with associated semidiscrete optimal state $\bar{y}_k \in Y_k$, analogously to Problem (**P**), follows by standard arguments. Likewise, we obtain first-order necessary and sufficient optimality conditions for $\bar{u}_k \in U_{\text{ad}}$ in the form

$$\hat{J}_k'(\bar{u}_k)(u - \bar{u}_k) = \langle \nu \bar{u}_k - B^* p_k(\bar{u}_k), u - \bar{u}_k \rangle_I \geq 0 \quad \forall u \in U_{\text{ad}}, \tag{12}$$

where $p_k = p_k(u) \in Y_k$ is the semidiscrete adjoint state, i.e. the solution of the semidiscrete adjoint equation

$$\mathbf{b}_k(\varphi_k, p_k) = (\varphi_k, y_d - y_k(u))_I \qquad \forall \varphi_k \in Y_k. \tag{13}$$

Note that the stability results from Theorem 1 are applicable to (13). By making use of the projection formula

$$\bar{u}_k = P_{\text{ad}} \left(\frac{1}{\nu} B^* p_k(\bar{u}_k) \right) \tag{14}$$

on this level of discretization we readily obtain a structural result for the semidiscrete optimal control:

Corollary 1. *From the projection formula* (14) *we deduce that all components of the optimal control $\bar{u}_k$ are piecewise constant in time, i.e. $\bar{u}_k \in U_k \cap U_{\text{ad}}$.*

Note that from Corollary 1, it is clear that $\bar{u}_k$ also solves the problem

$$\text{Minimize } \hat{J}_k(u_k) \text{ subject to } u_k \in U_k \cap U_{\text{ad}},$$

and thus the time discretization of the controls does not have to be discussed explicitely.

3.2 Discretization in space

Now, we introduce the spatial discretization of the optimal control problem, still in the spirit of e.g. [24]. We consider two- or three-dimensional shape regular meshes; see, e.g., [3], consisting of quadrilateral or hexahedral cells K, which constitute a nonoverlapping cover of the computational domain Ω. For $n = 1$, the cells reduce to subintervalls of Ω. We denote the mesh by $\mathscr{T}_h = \{K\}$ and define the discretization parameter h as a cellwise constant function by setting $h_{|_K} = h_K$ with the diameter h_K of the cell K. We use the symbol h also for the maximal cell size, i.e., $h = \max h_K$. On the mesh $\mathscr{T}_h$ we construct a conforming finite element space $V_h \subset V$ in the standard way

$$V_h = \{v \in V \,|\, v_{|_K} \in \mathscr{Q}(K) \text{ for } K \in \mathscr{T}_h\},$$

with basis $\{\Phi_h^j\}_{j=1,\dots,N}$, where $\mathscr{Q}(K)$ consists of shape functions obtained via bilinear transformations of polynomials up to degree one defined on a reference cell $\hat{K}$; cf. also Sec. 3.2 in [23]. Then, the space-time discrete finite element space

$$Y_{kh} = \{v_{kh} \in L^2(I, V_h) \mid v_{kh|I_m} \in \mathscr{P}_0(I_m, V_h),\ m = 1, 2, \dots, M\} \subset Y_k.$$

leads to a discrete version of the bilinear form $\mathbf{b}_k(\cdot,\cdot)$ for $y_{kh}, \varphi_{kh} \in Y_{kh}$, given by

$$\mathbf{b}_{kh}(y_{kh}, \varphi_{kh}) = (\nabla y_{kh}, \nabla \varphi_{kh})_I + \sum_{m=2}^{M} (y_{kh,m} - y_{kh,m-1}, \varphi_{kh,m}) + (y_{kh,1}, \varphi_{kh,1}).$$

Eventually, we obtain the so-called dG(0)cG(1) discretization of the state equation for given control $u \in U$: Find a state $y_{kh} = y_{kh}(u) \in Y_{kh}$ such that

$$\mathbf{b}_{kh}(y_{kh}, \varphi_{kh}) = (Bu, \varphi_{kh})_I + (f, \varphi_{kh})_I + (y_0, \varphi_{kh,1}) \qquad \forall \varphi_{kh} \in Y_{kh}. \tag{15}$$

Theorem 2. *Let Ass. 2.1 be satisfied. Then, for each $u \in U$ and possibly nonhomogneous initial condition $y_0 \in V$ and source term $f \in L^2(\Omega)$, there exists a unique solution $y_{kh} \in Y_{kh}$ of equation* (15) *satisfying the stability estimate*

$$\begin{aligned} &\|y_{kh}\|_I^2 + \|\nabla y_{kh}\|_I^2 + \|\Delta_h y_{kh}\|_I^2 + \sum_{i=1}^{M} k_m^{-1} \|y_{kh,i} - y_{kh,i-1}\|^2 \\ &\le C \left(|u|_I^2 + \|f\|_I^2 + \|\nabla \Pi_h y_0\|^2 \right) \end{aligned}$$

with a constant $C > 0$ independent of the discretization parameters k and h.

Proof. Again, this follows from [23], see Theorem 4.6, with the obvious modifications due to the structure of the control space.

Here, $\Delta_h \colon V_h \to V_h$ is defined by

$$(\Delta_h v_h, \varphi_h) = -(\nabla v_h, \nabla \varphi_h) \quad \forall \varphi_h \in V_h.$$

Repeating the steps from the semidiscrete setting leads to the introduction of the discrete control-to-state operator $S_{kh}\colon U \to Y_{kh}$, $y_{kh} = S_{kh}(u)$, the discrete reduced objective function $J_{kh} : U_{\mathrm{ad}} \to \mathbb{R}_0^+, \quad u \mapsto J(u, S_{kh}(u))$, and the discrete problem formulation

$$\text{Minimize } \hat{J}_{kh}(u) \text{ subject to } u \in U_{\mathrm{ad}}, \tag{$\mathbf{P}_{kh}$}$$

which, again by standard arguments, admits a unique optimal solution $\bar{u}_{kh} \in U_{\mathrm{ad}}$. First-order necessary and sufficient optimality condition for $\bar{u}_{kh} \in U_{\mathrm{ad}}$ are given by

$$\hat{J}'_{kh}(\bar{u}_{kh})(u - \bar{u}_{kh}) = \langle \nu \bar{u}_{kh} - B^* p_{kh}(\bar{u}_{kh}), u - \bar{u}_{kh} \rangle_I \geq 0 \quad \forall u \in U_{\mathrm{ad}}, \tag{16}$$

via the discrete adjoint state $p_{kh} \in Y_{kh}$ being the solution of the discrete adjoint equation

$$\mathbf{b}_{kh}(\varphi_{kh}, p_{kh}) = (\varphi_{kh}, y_d - y_{kh})_I \qquad \forall \varphi_{kh} \in Y_{kh}. \tag{17}$$

Let us conclude that the structure of (16), which is analogous to the continuous problem due to Galerkin discretization, is of central importance to adapt the error estimation techniques from [29], see Sec. 4.

3.3 A-priori Error Estimate

Let us end this section by stating an a-priori discretization error estimate between the optimal control $\bar{u}$ of (**P**) and the fully discrete solution $\bar{u}_{kh}$ of ($\mathbf{P}_{kh}$). The following theorem is a direct consequence of Theorem 6.1 in [23], where error estimates are proven for control functions $u \in L^2(I \times \Omega)$. The specific setting with finitely many time-dependent controls is considered for nonconvex problems with semilinear state equations in [27], Prop. 5.4.

Theorem 3. *Let $\bar{u}$ be the optimal control of Problem* (**P**) *and $\bar{u}_{kh}$ be the optimal control of Problem* ($\mathbf{P}_{kh}$)*. Then there exists a constant $C > 0$ independent of k and h, such that the following error estimate is satisfied:*

$$|\bar{u} - \bar{u}_{kh}|_I \leq C(k + h^2)$$

4 A-posteriori error analysis for an approximate solution to ($\mathbf{P}_{kh}$)

If the possibly high-dimensional optimization problem ($\mathbf{P}_{kh}$) is not solved directly, but an approximate solution $\bar{u}^p_{kh} \in U_k \cap U_{\mathrm{ad}}$ is obtained by e.g. a POD Galerkin approximation, see Sec. 5, one is interested in estimating the error

$$\varepsilon^p_{kh} := |\bar{u}^p_{kh} - \bar{u}_{kh}|_I,$$

without knowing $\bar{u}_{kh}$. Together with an available error estimate for

$$\varepsilon_{kh} := |\bar{u} - \bar{u}_{kh}|_I,$$

this leads to an estimate for the error between the real optimal solution $\bar{u}$ and the computed solution $\bar{u}^p_{kh}$,

$$\varepsilon := |\bar{u} - \bar{u}^p_{kh}|_I \leq \varepsilon_{kh} + \varepsilon^p_{kh},$$

where the influence of the discretization error on the one hand and the model reduction error on the other hand are clearly separated. We will use Theorem 3 for the first part. We point out that ε_{kh} may also be estimated by other available e.g. a-posteriori error estimators, and now focus on developing an estimate for the second part. Since the discretization has been obtained by a Galerkin method, and optimality conditions from Sec. 3 are available, we can apply the arguments and techniques from [29]. Utilizing the notation $y^p_{kh} = S_{kh}(u^p_{kh})$ as well as $p^p_{kh} = p_{kh}(u^p_{kh})$, we obtain the following:

Lemma 3. *Let $u^p_{kh} \in U_k \cap U_{\mathrm{ad}}$ be arbitrary. Define a function $\xi^p_k \in U_k$ componentwise by*

$$(\xi^p_k)_i := \left(\nu u^p_{kh} - B^* p^p_{kh}\right)_i, \quad i = 1, \ldots, N_u,$$

as well as the index sets of active constraints

$$\mathscr{I}_i^- := \{1 \leq m \leq M \mid (u^p_{kh})_{i|_{I_m}} = u_{a,i}\}, \quad i = 1, \ldots, N_u,$$
$$\mathscr{I}_i^+ := \{1 \leq m \leq M \mid (u^p_{kh})_{i|_{I_m}} = u_{b,i}\}, \quad i = 1, \ldots, N_u,$$

the active sets $\mathscr{A}_i^- := \bigcup\{I_m \mid m \in \mathscr{I}_i^-\}$, $\mathscr{A}_i^+ := \bigcup\{I_m \mid m \in \mathscr{I}_i^+\}$ and the inactive set $\mathscr{A}_i^\circ := I \setminus (\mathscr{A}_i^- \cup \mathscr{A}_i^+)$. Then the function $\zeta_k \in U_k$ defined componentwise by

$$\zeta_{k,i} = [\xi^p_{k,i}]_- = -\min\{0, \xi^p_{k,i}\} \text{ on } \mathscr{A}_i^-,$$
$$\zeta_{k,i} = -[\xi^p_{k,i}]_+ = -\max\{0, \xi^p_{k,i}\} \text{ on } \mathscr{A}_i^+,$$
$$\zeta_{k,i} = -\xi^p_{k,i} \text{ on } \mathscr{A}_i^\circ$$

for $i = 1, \ldots, N_u$ satisfies the perturbed variational inequality

$$\langle \nu u^p_{kh} - B^* p^p_{kh} + \zeta_k, u - u^p_{kh} \rangle_I \geq 0 \qquad \forall u \in U_k \cap U_{\mathrm{ad}}.$$

Proof. Note first that due to the piecewise constant time discretization, the function ζ_k is an element of U_k. Now, direct calculations for $u \in U_k \cap U_{\mathrm{ad}}$ shows:

$$\langle \nu u_{kh}^p - B^* p_{kh}^p + \zeta_k, u - u_{kh}^p \rangle = \sum_{i=1}^{N_u} \sum_{m=1}^{M} \int_{I_m} (\xi_k^p + \zeta_k)_i (u - u_{kh}^p)_i \, dt$$

$$= \sum_{i=1}^{N_u} \sum_{m \in \mathscr{I}_i^-} k_m (\xi_{k,m}^p + [\xi_{k,m}^p]_-)_i (u - u_a)_i + \sum_{m \in \mathscr{I}_i^+} k_m (\xi_{k,m}^p - [\xi_{k,m}^p]_+)_i (u - u_b)_i \geq 0,$$

where we have used that $(u-u_a)_i \geq 0$ and $(u-u_b)_i \leq 0$.

Theorem 4. *Let $\bar{u}_{kh}$ be the optimal solution to $(\mathbf{P}_{kh})$ with associated state $\bar{y}_{kh}$ and adjoint state $\bar{p}_{kh}$. Suppose that $u_{kh}^p \in U_k \cap U_{\mathrm{ad}}$ is chosen arbitrarily with associated state $y_{kh}^p = y_{kh}(u_{kh}^p) \in Y_{kh}$ and adjoint state $p_{kh}^p = p_{kh}(u_{kh}^p) \in Y_{kh}$, and let $\zeta_k \in U_k$ be given as in Lemma 3. Then the following estimate is satisfied:*

$$|\bar{u}_{kh} - u_{kh}^p|_I \leq \frac{1}{\nu} |\zeta_k|_I.$$

Proof. The variational inequality (15)-(17) and Lemma 3 imply

$$\begin{aligned}
0 &\leq \langle \nu \bar{u}_{kh} - B^* \bar{p}_{kh}, u_{kh}^p - \bar{u}_{kh} \rangle_I + \nu \langle u_{kh}^p - B^* p_{kh}^p + \zeta_k, \bar{u}_{kh} - u_{kh}^p \rangle_I \\
&= -\nu |\bar{u}_{kh} - u_{kh}^p|_I^2 + \langle B^*(p_{kh}^p - \bar{p}_{kh}), u_{kh}^p - \bar{u}_{kh} \rangle_I - \langle \zeta_k, u_{kh}^p - \bar{u}_{kh} \rangle_I \\
&= -\nu |\bar{u}_{kh} - u_{kh}^p|_I^2 + (p_{kh}^p - \bar{p}_{kh}, B(u_{kh}^p - \bar{u}_{kh}))_I - \langle \zeta_k, u_{kh}^p - \bar{u}_{kh} \rangle_I \\
&= -\nu |\bar{u}_{kh} - u_{kh}^p|_I^2 + (B(u_{kh}^p - \bar{u}_{kh}), p_{kh}^p - \bar{p}_{kh})_I - \langle \zeta_k, u_{kh}^p - \bar{u}_{kh} \rangle_I \\
&= -\nu |\bar{u}_{kh} - u_{kh}^p|_I^2 + \mathbf{b}(y_{kh}^p - \bar{y}_{kh}, p_{kh}^p - \bar{p}_{kh}) - \langle \zeta_k, u_{kh}^p - \bar{u}_{kh} \rangle_I \\
&= -\nu |\bar{u}_{kh} - u_{kh}^p|_I^2 - \|y_{kh}^p - \bar{y}_{kh}\|_I^2 - \langle \zeta_k, u_{kh}^p - \bar{u}_{kh} \rangle_I.
\end{aligned}$$

From this calculation, we conclude that

$$\nu |\bar{u}_{kh} - u_{kh}^p|_I^2 \leq |\zeta_k|_I |u_{kh}^p - \bar{u}_{kh}|_I, \qquad \Longrightarrow \qquad |\bar{u}_{kh} - u_{kh}^p|_I \leq \frac{1}{\nu} |\zeta_k|_I.$$

Combining the results of Theorem 3 and Theorem 4, we directly obtain

Corollary 2. *Let $\bar{u} \in U_{\mathrm{ad}}$ be the optimal control of Problem $(\mathbf{P})$, let $u_{kh}^p \in U_k \cap U_{\mathrm{ad}}$ be chosen arbitrarily, and let $\zeta_k \in U_k$ be given as in Lemma 3. Then there exists a constant $C > 0$ independent of k and h, such that the following error estimate is fulfilled:*

$$|\bar{u} - u_{kh}^p|_I \leq C(k + h^2) + \frac{1}{\nu} |\zeta_k|_I.$$

5 The POD Galerkin discretization

In this section, we construct a problem specific subspace $V_h^\ell \subseteq V_h$ with significantly smaller dimension $\dim V_h^\ell = \ell \ll N = \dim V_h$ such that the projection of an element

y_{kh} on the reduced state space $Y_{kh}^\ell = \{\phi_{kh} \in Y_{kh} \mid \forall m = 1,...,M : \phi_{kh}|_{I_m} \in \mathscr{P}_0(I_m, V_h^\ell)\}$ is still a good approximation of y_{kh}. More precisely, for a given basis rank ℓ, we choose orthonormal ansatz functions $\psi_h^1,...,\psi_h^\ell \in V_h$ such that $y_{kh} - \mathscr{P}_h^\ell y_{kh}$ is minimized with respect to $\|\cdot\|_{L^2(I,V_h)}$ where $\mathscr{P}_h^\ell : V_h \to V_h^\ell = \operatorname{span}(\psi_h^1,...,\psi_h^\ell)$ denotes the canonical projector $\mathscr{P}_h^\ell(y_{kh}(t)) = \sum_{l=1}^\ell \langle y_{kh}(t), \psi_h^l\rangle_{V_h} \psi_h^l$. Hence, these basis functions $\psi_h^1,...,\psi_h^\ell$ are given as a solution to the optimization problem

$$\min_{\psi_h^1,...,\psi_h^\ell \in V_h} \int_0^T \|y_{kh}(t) - \mathscr{P}_h^\ell y_{kh}(t)\|_{V_h}^2 \,\mathrm{d}t \qquad \text{subject to} \qquad \langle \psi_h^i, \psi_h^j\rangle_{V_h} = \delta_{ij}, \quad (18)$$

where δ_{ij} denotes the Kronecker delta. Since the integrand is piecewise constant on the time intervals $I_1,...,I_m$, we replace (18) by

$$\min_{\psi_h^1,...,\psi_h^\ell \in V_h} \sum_{m=1}^M k_m \|y_{kh,m} - \mathscr{P}_h^\ell y_{kh,m}\|_{V_h}^2 \qquad \text{subject to} \qquad \langle \psi_h^i, \psi_h^j\rangle_{V_h} = \delta_{ij}. \quad (19)$$

Due to the discrete structure of the spaces Y_{kh} and V_h, problem (18) is further equivalent to a finite dimensional linear system: Let $(\mathrm{A}_h)_{ij} = \langle \phi_h^i, \phi_h^j\rangle_{V_h} \in \mathbb{R}^{N\times N}$ denote the stiffness matrix of the finite elements sample $(\phi_h^1,...,\phi_h^N)$ and let $\Theta_k \in \mathbb{R}^{M\times M}$ be the diagonal matrix consisting of the time weights $k_1,...,k_M$. Further, let $\mathrm{Y}_{kh} \in \mathbb{R}^{N\times M}$ be the coefficient matrix of $y_{kh} \in Y_{kh}$ such that $y_{kh}(t_j) = \sum_{i=1}^N \mathrm{Y}_{kh}^{ij} \phi_h^i$ holds. Denoting the i-th column of a matrix by the superscript $^{\cdot,i}$, we get the fully discretized problem

$$\min_{\psi_{h\ell} \in \mathbb{R}^{N\times\ell}} \sum_{m=1}^M \Theta_k^{mm} ((\mathrm{Y}_{kh}^{\cdot,m})^\mathrm{T} - (\mathrm{Y}_{kh}^{\cdot,m})^\mathrm{T} \mathrm{A}_h \psi_{h\ell} \psi_{h\ell}^\mathrm{T}) \mathrm{A}_h (\mathrm{Y}_{kh}^{\cdot,m} - \psi_{h\ell} \psi_{h\ell}^\mathrm{T} \mathrm{A}_h \mathrm{Y}_{kh}^{\cdot,m}) \quad (20)$$
$$\text{subject to} \qquad \psi_{h\ell}^\mathrm{T} \mathrm{A}_h \psi_{h\ell} = \mathrm{Id}(\ell).$$

The solution matrix $\psi_{h\ell} \in \mathbb{R}^{N\times\ell}$ to (20) provides the coefficients of the optimal basis elements $\psi_h^1,...,\psi_h^\ell$ to (18): $\psi_h^l = \sum_{i=1}^N \psi_{h\ell}^{il} \phi_h^i$.

A solution to problem (18) is called a *rank-ℓ POD basis* to the trajectory $y_{kh} \in Y_{kh}$ and can be determined by solving the corresponding eigenvalue problem

$$\mathscr{R}(y_{kh})\psi_h = \lambda \psi_h, \qquad \mathscr{R}(y_{kh}) = \int_0^T \langle \cdot, y_{kh}(t)\rangle_{V_h} y_{kh}(t) \,\mathrm{d}t, \quad (21)$$

choosing $\psi_h^1,...,\psi_h^\ell \in V_h$ as the eigenfunctions of $\mathscr{R}_{kh} := \mathscr{R}(y_{kh}) : V_h \to V_h$ corresponding to the ℓ largest eigenvalues $\lambda_1 \geq \cdots \geq \lambda_\ell > 0$, cf. [9], Sec. 2.2. Notice that the adjoint operator $\mathscr{R}_{kh}^*$ is compact due to the Hilbert-Schmidt theorem, i.e. $\mathscr{R}_{kh}$ is compact as well, and that $\mathscr{R}_{kh}$ is non-negative, so a complete decomposition of V_h into eigenfunctions of $\mathscr{R}_{kh}$ is available, and each eigenvalue except for possibly zero has finite multiplicity, see [9], Lemma 2.12 and Theorem 2.13.

$\mathscr{R}_{kh}$ has the semidiscrete representation

$$\mathscr{R}_{kh} = \sum_{m=1}^{M} k_m \langle \cdot, y_{kh,m} \rangle_{V_h} y_{kh,m}.$$

Solving the eigenvalue problem (21) for the eigenfunctions $\psi_h^1, ..., \psi_h^\ell \in V_h$ coincides with the determination of their coefficient matrix $\psi^\ell \in \mathbb{R}^{N \times \ell}$. Since

$$\mathscr{R}_{kh} \psi_h^l = \sum_{m=1}^{M} k_m \mathrm{Y}_{kh}^{\cdot,m} (\mathrm{Y}_{kh}^{\cdot,m})^{\mathrm{T}} \mathrm{A}_h \psi_{h\ell}^{\cdot,l} = \mathrm{Y}_{kh} \Theta_k \mathrm{Y}_{kh}^{\mathrm{T}} \mathrm{A}_h \psi_{h\ell}^{\cdot,l},$$

the columns of $\psi_{h\ell}$ are given as the solution to the discretized eigenproblem

$$R_{kh} \psi_h = \lambda \psi_h, \qquad R_{kh} = \mathrm{Y}_{kh} \Theta_k \mathrm{Y}_{kh}^{\mathrm{T}} \mathrm{A}_h \tag{22}$$

corresponding to the discretized problem (20).

In practice, one may replace the matrix R_{kh} in (22) by the symmetrized one

$$\tilde{R}_{kh} = \sqrt{\mathrm{A}_h} \mathrm{Y}_{kh} \Theta_k \mathrm{Y}_{kh}^{\mathrm{T}} \sqrt{\mathrm{A}_h} \in \mathbb{R}^{N \times N},$$

calculate the eigenvectors $\tilde{\psi}_{h\ell}^1, ..., \tilde{\psi}_{h\ell}^\ell \in \mathbb{R}^N$ and gain $\psi_{h\ell}^{\cdot,l}$ by the transformation

$$\tilde{\psi}_{h\ell}^l = \sqrt{\mathrm{A}_h} \psi_{h\ell}^{\cdot,l}, \qquad l = 1, ..., \ell;$$

in addition to the effort of solving the eigenvalue problem itself, the root of the mass matrix has to be calculated here and ℓ linear systems of dimension N have to be solved.

Depending on the spatial dimension of the control problem and on the number of gridpoints for the time and space discretizations, it may be convenient to provide the M-dimensional eigenvalue decomposition for the adjoint operator $\mathscr{K}_{kh} = \mathscr{R}_{kh}^*$ instead. If $M < N$ holds, we deal with eigenvector-eigenvalue pairs $(\tilde{\psi}_{h\ell}^l, \lambda^l)$ of the matrix

$$\tilde{K}_{kh} = \sqrt{\Theta_k} \mathrm{Y}_{kh}^{\mathrm{T}} \mathrm{A}_h \mathrm{Y}_{kh} \sqrt{\Theta_k} \in \mathbb{R}^{M \times M}$$

instead. The POD basis coefficient vectors then are given by the easier transformation

$$\psi_{h\ell}^{\cdot,l} = \frac{1}{\lambda^l} \mathrm{Y}_{kh} \sqrt{\Theta_k} \tilde{\psi}_{h\ell}^l, \qquad l = 1, ..., \ell.$$

In this case, neither a transformation of the eigenvectors which would require an additional solving step of a linear system is required nor a matrix root has to be determined; the root of the time weights matrix Θ_k is given elementwise since this matrix is diagonal.

The following a-priori error estimate holds, cf. Proposition 3.3 in [15]:

$$\int_0^T \|y_{kh}(t) - \mathscr{P}_h^\ell y_{kh}(t)\|_{V_h}^2 \,\mathrm{d}t = \sum_{m=1}^M \left\| y_{kh,m} - \mathscr{P}_h^\ell y_{kh,m} \right\|_{V_h}^2 = \sum_{l=\ell+1}^{\operatorname{rank}(\mathscr{R}_{kh})} \lambda^l.$$

After a POD basis of some reference trajectory $y_{kh} \in Y_{kh}$ is constructed, we introduce the *reduced state space*

$$Y_{kh}^\ell = \{\phi_{kh} \in Y_{kh} \mid \forall m = 1,\dots,M : \phi_{kh}|_{I_m} \in \mathscr{P}_0(I_m, V_h^\ell)\}$$

and consider the reduced-order optimization problem

$$\min_{(y_{kh}^\ell, u) \in Y_{kh}^\ell \times U_{\mathrm{ad}}} \frac{1}{2} \sum_{m=1}^M k_m \|y_{kh,m} - y_{d,m}\|_{V_h}^2 + \frac{\nu}{2}|u|_I^2. \tag{23}$$

The first-order optimality conditions of (23), consisting of a reduced state equation corresponding to (4), a reduced adjoint state equation corresponding to (9) and a control equation corresponding to (10), read as

$$\mathbf{b}_{kh}(y_{kh}, \varphi_{kh}) - (Bu_{kh}, \varphi_{kh})_I = 0 \qquad \forall \varphi_{kh} \in Y_{kh}^\ell, \tag{24a}$$

$$\mathbf{b}_{kh}(\varphi_{kh}, p_{kh}) - (\varphi_{kh}, y_d - y_{kh})_I = 0 \qquad \forall \varphi_{kh} \in Y_{kh}^\ell, \tag{24b}$$

$$\langle \nu u_{kh} - B^* p_{kh}, u - u_{kh} \rangle_I \geq 0 \qquad \forall u \in U_{\mathrm{ad}}. \tag{24c}$$

Since the *optimal* trajectory $\bar{y}_{kh}$ is not available in practice to build up an appropriate POD basis for the reduced-order model, different methods have been developed recently on how to construct a reference trajectory $\tilde{y}_{kh}$ which covers enough dynamics of $\bar{y}_{kh}$ to build up an accurate reduced order space V_h^ℓ.

If the desired state y_d is smooth and the regularization parameter ν is sufficiently large, the dynamics of $\bar{y}_{kh}$ are usually simple enough such that the state solution $\tilde{y}_{kh}$ to some more or less arbitrary reference control such as $\tilde{u} \equiv 1$ generates a suitable reduced space V_h^ℓ. Otherwise, if $\tilde{y}_{kh}$ differs too much from the optimal solution $\bar{y}_{kh}$, it may be necessary to choose an inefficiently large basis rank ℓ to represent the more complex dynamics of $\bar{y}_{kh}$ in the eigenfunctions of $\tilde{y}_{kh}$. Moreover, though the properties of $\mathscr{R}(\tilde{y}_{kh})$ should guarantee a suitable approximation of $\bar{y}_{kh}$ in V_h^ℓ if ℓ is sufficiently large, numerical instabilities arise especially if λ_ℓ comes close to zero: In this case, the set of eigenfunctions to $\mathscr{R}(\tilde{y}_{kh})$ corresponding to the nonzero eigenvalues is not enlarged by a basis of $\operatorname{range}(\mathscr{R})^\perp$; instead, numerical noise is added on the eigenfunctions so that the POD basis is not improved, but even perturbed, cf. Fig. 4. Consequently, the system matrices of the reduced-order model become singular and the reduced order solutions get instable. One way to balance out this problem is to provide a *basis update* which improves the reference control $\tilde{y}_{kh}$ and hence the ansatz space V_h^ℓ.

Another method to provide accurate reduced models is the so-called *Optimality System POD*, where the (high-dimensional) conditions for the optimal snapshot sample (15), (21) are included into the reduced optimization problem. The resulting first-order optimality system which by construction includes the optimal dynamics of the state is solved by a less accurate, but cheap routine for the high-order com-

ponents and by a more complex method providing higher convergence orders for the reduced equations, see [17], for instance. Let $(\psi_h^1, ..., \psi_h^\ell) \subseteq V_h$ be a rank-ℓ POD basis computed from some reference state $\tilde{y}_{kh}$ and let u_{kh}^ℓ, $\ell = 1, 2, ...$, denote the control solution to the reduced-order optimality system (24). We study the development of the control errors $\varepsilon_{\text{ex}}^\ell = |\bar{u} - \bar{u}_{kh}^\ell|_I$ for various ℓ. A stagnation of (the order of) $\varepsilon_{\text{ex}}^\ell$ may be caused by three different effects:

1. The chosen basis ranks are still too small to represent the corresponding optimal state solutions $\bar{y}_{kh}$ in an appropriate way. Adding some more basis vectors may finally lead to a decay of $\varepsilon_{\text{ex}}^\ell$; the stagnation is not necessarily an indication for a badly chosen reference trajectory. Indeed, even the *optimal* POD basis corresponding to the exact FE solution $\bar{y}_{kh}$ does not guarantee small errors $\varepsilon_{\text{ex}}^\ell$ for small basis ranks, cf. our numerical tests in the next section.
2. The vector space spanned by the eigenfunctions of $\mathscr{R}(y_{kh})$ is exploited before $\varepsilon_{\text{ex}}^\ell$ decays below the desired exactness ε. As mentioned above, additional POD elements will not improve the error decay in this case. Further, the available information may not be sufficient to extend the current basis by additional vectors at all.
3. The accuracy of the finite element model $\varepsilon_{kh}^{\text{ex}} = |\bar{u} - \bar{u}_{kh}|_I$ is reached. In this case, expanding the POD basis may decrease the error $\varepsilon_{kh}^\ell = |\bar{u}_{kh} - \bar{u}_{kh}^\ell|_I$ between the high-dimensional and the low-dimensional approximation of $\bar{u}$, but not the actually relevant error $\varepsilon_{\text{ex}}^\ell$ between $\bar{u}_{kh}^\ell$ and the exact control solution $\bar{u}$.

Algorithm 1 (Reduced Order Modeling)

Require: Basis ranks $\ell_{\min} < \ell_{\max}$, initial POD basis elements $\psi_h^1, ..., \psi_h^{\ell_{\max}} \in V_h$, maximal number of basis updates $j_{\max}$.

Estimate finite element error $\varepsilon_{kh}^{\text{ex}} = |\bar{u} - \bar{u}_{kh}|_I$. Set $j = 1$, $\ell = \ell_{\min}$.
while $j \leq j_{\max}$ **do**
  Set $\psi_h^\ell = (\psi_1, ..., \psi_\ell)$.
  Calculate optimal control $\bar{u}_{kh}^\ell$ to the ℓ-dimensional reduced-order model.
  Estimate ROM residual $\varepsilon_{kh}^\ell = |\bar{u}_{kh} - \bar{u}_{kh}^\ell|_I$.
  if $\varepsilon_{kh}^\ell \leq \varepsilon_{kh}^{\text{ex}}$ **then**
    break (optimal accuracy reached)
  else if $\ell < \ell_{\max}$ **then**
    Set $\ell = \ell + 1$. (enlarge POD basis)
  else
    Calculate new POD basis elements $\psi_h^1, ..., \psi_h^{\ell_{\max}} \in V_h$. (update POD basis)
    Set $\ell = \ell_{\min}$ and $j = j + 1$.
  end if
end while

We propose to choose a minimal and a maximal basis rank $\ell_{\min}$, $\ell_{\max}$ at the beginning and to increase the reduced model rank ℓ frequently, starting from $\ell_{\min}$ on, until $\varepsilon_{\text{ex}}^\ell$ decays below the desired exactness ε or $\ell_{\max}$ is reached; in the latter case, a basis update is provided, choosing the lastly determined POD optimal control $\bar{u}_{kh}^{\ell_{\max}}$ to calculate new snapshots and resetting the model rank on $\ell_{\min}$.

6 Numerical Results

We test our findings combined in Algorithm 1 with the aid of an analytical test problem where the exact control, state and adjoint solutions $\bar{u}, \bar{y}, \bar{p}$ are known explicitely. For this purpose, we choose the one-dimensional spatial domain $\Omega = (0, 2\pi)$, the control space $U = L^2(I, \mathbb{R}^1)$ consisting of a single-component control $u = u(t)$ on the time interval $I = [0, \frac{\pi}{2}]$, the single shape function $\chi(x) = \sin(x)$, the lower and upper control bounds $u_a = -5$, $u_b = 5$ and the regularization parameter $\nu = 1$. To realize the optimal triple

$$\bar{y}(t,x) = \sin(x)\cos(x\exp(t)), \qquad \bar{p}(t,x) = \sin(x)\cos(t), \tag{25a}$$

$$\bar{u}(t) = P_{\mathrm{ad}}\left(\left\{\frac{\pi}{\nu}\cos(t)\right\} + \left\{10\sin(\exp(2t))\right\}\right), \tag{25b}$$

we define the source term $f \in L^2(I \times \Omega)$, the initial value $y_0 \in H_0^1(\Omega)$ and the desired state $y_d \in L^2(I \times \Omega)$ by

$$\begin{aligned} f(t,x) = &-\sin(x)\sin(xe^t)xe^t + \sin(x)\cos(xe^t) + \cos(x)\sin(xe^t)e^t \\ &+ \cos(x)\sin(xe^t)e^t + \sin(x)\cos(xe^t)e^{2t} - \chi(x)\bar{u}(t), \\ y_0(x) = &\sin(x)\cos(x), \\ y_d(t,x) = &\sin(x)\sin(t) + \sin(x)\cos(t) + \bar{y}(t,x). \end{aligned}$$

Due to technical reasons, we introduce a desired control $u_d \in U$ in addition by

$$u_d(t) = 10\sin(\exp(2t))$$

and consider the adapted objective functional $\tilde{J}(y,u) = J(y, u - u_d)$. In the optimality system, this has no impact on the state equation or the adjoint equation; the adapted variational inequality for the control now reads as $\langle \nu(\bar{u} - u_d) - B^*\bar{p}, u - \bar{u}\rangle_I \geq 0$ for all $u \in U_{\mathrm{ad}}$. By direct recalculation one sees that the functions in (25) fulfill the adapted optimality equations. Fig. 1 illustrates the optimal solution (25).

Fig. 1 The optimal state $\bar{y}$ (left), the optimal control $B\bar{u}$ (center) and the optimal adjoint state $\bar{p}$ (right) of the test setting.

The full-order optimality system (15)-(17) as well as the reduced-order one (24) are solved by a simple fixpoint iteration $u_{n+1} = F(u_n) = P_{\mathrm{ad}}(\frac{1}{\nu}B^*p(y(u)))$ with admissible initial control u_0. This procedure generates a converging sequence $(u_{n+1})_{n\in\mathbb{N}} \subseteq U_{\mathrm{ad}}$ with limit $\bar{u}$ given that ν is not too small. In this case, F is a contracting selfmapping on U_{ad} and the Banach fixpoint theorem provides decay rates for the residual $|\bar{u} - u_n|_I$, cf. [9], Sec. 5.5. Compared to numerical strategies which

provide higher convergence rates such as Newton methods, the numerical effort within the single iterations is small since no coupled systems of PDEs have to be solved.

6.1 Finite element error estimation

In order to be able to combine the a-priori finite element error estimates from Sec. 3.3 with the a-posteriori error estimate for the POD approximation in a reasonable way, we need to estimate the constant appearing in Theorem 3. More precisely, we will estimate two constants $C_t, C_x > 0$ such that $\|\bar{u} - \bar{u}_{kh}\|_I \approx C_t k + C_x h^2$ holds. In this way, we receive slightly better results then with the constant C presented in Theorem 3; choose $C = \max(C_t, C_x)$ for convenience. The dependency between the time and space discretization quantities h, k and the resulting discretization errors is shown in Fig. 2 (left); the quality of this *error indicator* is shown in Fig. 2 (right).

Fig. 2 On the left, we show that the exact errors of time integration $k \mapsto \varepsilon^{\mathrm{ex}}_{kh_0}$ on a sufficiently fine spatial grid $h_0^2 \ll k$ and the exact spatial errors $h^2 \mapsto \varepsilon^{\mathrm{ex}}_{k_0 h}$ with sufficiently small time steps $k_0 \ll h^2$ evolve approximatively linear in logarithmic scales. On the right, on sees that $\mathrm{Ind}(k,h) \approx \varepsilon^{\mathrm{ex}}_{kh}$ holds given that the time and space grids are not too coarse: The bounds are sharp, but not rigorous.

We estimate such constants C_t, C_x by solving the discretized optimality system (15)-(17) on grids of different grid widths:

$$C_t = \frac{1}{k_1}|\bar{u}_{k_1 h_0} - \bar{u}_{k_2 h_0}|_I \ (h_0^2, k_2 \ll k_1), \quad C_x = \frac{1}{h_1^2}|\bar{u}_{k_0 h_1} - \bar{u}_{k_0 h_2}|_I \ (h_2^2, k_0 \ll h_1^2).$$

Notice that this procedure does not guarantee that $\mathrm{Ind}(k,h) = C_t k + C_x h^2$ provides an upper bound for the FE error. In our numerical example, we choose the parameters $h_0 = 3.14\text{e-}2$ and $k_1 = 3.08\text{e-}2$, $k_2 = 1.57\text{e-}3$ as well as $k_0 = 3.93\text{e-}3$ and $h_1 = 2.99\text{e-}1$, $h_2 = 6.22\text{e-}2$ and get $C_t = 0.2$, $C_x = 0.184$.

6.2 Model reduction error estimation

We devide the time interval into $M = 6400$ subintervals and the spatial domain into $N = 500$ subdomains. In this case, $k = 2.45\text{e-}4$ and $h = 1.26\text{e-}2$ hold. With the growth constants given above, we expect a finite element accuracy of the magnitude $\varepsilon^{\mathrm{ex}}_{kh} = 7.82\text{e-}5$. Let $\tilde{y}_{kh}$ be a perturbation of the optimal finite element solution $\bar{y}_{kh}$ such that $\|\bar{y}_{kh} - \tilde{y}_{kh}\|_I$ is of the order $\tilde{\varepsilon} = 1.00\text{e-}7$. Although $\tilde{\varepsilon} < \varepsilon^{\mathrm{ex}}_{kh}$ holds true, the POD error ε^{ℓ}_{kh} and hence also the exact error $\varepsilon^{\ell}_{\mathrm{ex}}$ between the controls $\bar{u}^{\ell}_{kh}$ and $\bar{u}_{kh}$ or

$\bar{u}$, respectively, do not reach the desired accuracy $\varepsilon_{kh}^{\mathrm{ex}}$ independent of the chosen basis rank ℓ, cf. Fig. 3 (left); the POD elements react very sensitive if the corresponding snapshots are covered by noise. Notice that a perturbation of the control generating the snapshots would not have this destabilizing effect on the POD basis.

After providing a basis update, we observe that the low-order model error ε_{kh}^{ℓ} decays far below the high-order model accuracy $\varepsilon_{kh}^{\mathrm{ex}}$ for increasing basis rank ℓ while the exact error $\varepsilon_{\mathrm{ex}}^{\ell}$ stagnates on the level of $\varepsilon_{kh}^{\mathrm{ex}}$ as expected, cf. Fig. 3 (right). Starting with $\ell_{\min} = 12$, the Algorithm stops successfully after three basis extensions without requiring a further basis update.

Fig. 3 Here we present the decay behavior of the errors ε_{kh}^{ℓ} and $\varepsilon_{\mathrm{ex}}^{\ell}$ in dependence of the chosen POD basis rank ℓ as well as the desired FE error level $\varepsilon_{kh}^{\mathrm{ex}}$. On the left, a POD basis belonging to a small pointwise perturbation of the optimal state is applied to bouild up the reduced order model. On the right, we use an updated POD basis.

In Fig. 4 we compare the two POD bases. It turns out that overall, the first fifteen basis elements coincide, except of possibly the sign. Then, the noise starts to dominate the perturbed basis; the seventeenth basis function has no structure any more (left). In contrast, the updated basis spans a subspace of V_h with approximative dimension 34. POD elements of higher rank order than 34 get unstable as well (right), but here, the spanned space is already sufficiently large and includes enough dynamics of the optimal state solution to represent $\bar{y}_{kh}$ up to FE precision.

Fig. 4 Some elements of the perturbed (left) and the updated (right) POD basis in comparison.

Finally, we compare the effort of the full-order model with the reduced one. The calculation times of the single steps are presented in Tab. 1; the calculations are provided on an Intel(R) Core(Tm) i5 2.40 GHz processor.

Process (ROM)	**Time**	**#**	**Total**
Estimate FE error	6.04 sec	1×	6.04 sec
Calculate snapshots	3.17 sec	2×	6.34 sec
Solve eigenvalue problem	19.09 sec	2×	38.17 sec
Assemble reduced system	0.33 sec	2×	0.66 sec
Solve reduced system	0.87 sec	50×	43.65 sec
Evaluate error estimator	7.28 sec	2×	14.56 sec
Total			109.40 sec

Process (FEM)	**Time**	**#**	**Total**
Solve full-order system	25.81 sec	30×	774.27 sec
Total			774.27 sec

Table 1 The duration of the single processes within the reduced order modelling routine with adaptive basis selection.

We use $N = 4000$ finite lements now and $M = 4000$ time steps. Without model reduction, 30 fixpoint iterations are required, taking 774.27 seconds in total. To avoid noise in the POD elements, we choose adaptive basis ranks $\ell_{\min} = \ell_{\max} = \min\{20, \ell_\sigma\}$ where $\ell_\sigma = \max\{\ell \mid \lambda_\ell > 1.0\text{e-}12\}$: No basis elements corresponding to eigenvalues close to zero shall be used to build up the reduced model. As before, the first solving of the reduced optimality system requires 30 iterations, but the model error ε_{kh}^ℓ is still above the desired accuracy $\varepsilon_{kh}^{\text{ex}}$. The rank of the reduced-order problem is 14 (since $\lambda_{15} = 4.03\text{e-}13$). Providing one basis update with the snapshots $y_{kh}^\ell(\bar{u}_{kh}^{14})$, the new POD elements include more dynamics of the problem although the eigenvalues decay as before; now, we have $\ell = 15$ (since $\lambda_{16} = 2.61\text{e-}13$), the fixpoint routine terminates after 20 iterations and ε_{kh}^ℓ decays below the FE accuracy. The reduced-order modelling takes 109.40 seconds, 14.09% of the full-order solving.

Acknowledgements Ira Neitzel is grateful for the support of her former host institution Technische Universität München. Martin Gubisch gratefully acknowledges support by the German Science Fund DFG grant VO 1658/2-1 *A-posteriori-POD Error Estimators for Nonlinear Optimal Control Problems governed by Partial Differential Equations*.

References

1. Bergam, A., Bernardi, C., Mghazli, Z.: A posteriori analysis of the finite element discretization of some parabolic equations. Math. Comp. **74**(251), 1117–1138 (2004)
2. Chrysafinos, K.: Convergence of discontinuous Galerkin approximations of an optimal control problem associated to semilinear parabolic PDEs. ESAIM M2AN **44**(1), 189–206 (2010)
3. Ciarlet, P.G.: The Finite Element Method for Elliptic Problems. North-Holland, Amsterdam (1987),

4. Dautray, R., Lions, J.L.: Evolution Problems. Mathematical Analysis and Numerical Methods for Science and Technology **5**, Springer Berlin Heidelberg (2000)
5. Eriksson, K., Johnson, C., Thomée, V.: Time discretization of parabolic problems by the discontinuous Galerkin method. ESAIM Math. Model. Num. **19**, 611–643 (1985)
6. Eriksson, K., Estep, D., Hansbo, P., Johnson, C.: Computational Differential Equations Cambridge University Press, UK (1996)
7. Evans, L.C.: Partial Differential Equations. American Math. Society, Providence, Rhode Island (1998)
8. Gong, W., Hinze, M.: Error estimates for parabolic optimal control problems with control and state constraints. Comput. Optim. Appl. **56**(1), 131–151 (2013)
9. Gubisch, M., Volkwein, S.: POD a-posteriori error analysis for optimal control problems with mixed control-state constraints. Comp. Opt. Appl. **58**(3), 619–644 (2014)
10. Gubisch, M., Volkwein, S.: Proper orthogonal decomposition for linear-quadratic optimal control. to appear in: P. Benner, A. Cohen, M. Ohlberger, and K. Willcox (eds.): Model Reduction and Approximation: Theory and Algorithms. SIAM, Philadelphia, PA, 2016. preprint available: http://nbn-resolving.de/urn:nbn:de:bsz:352-250378
11. Hinze, M.: A variational discretization concept in control constrained optimization: the linear-quadratic case. Comput. Optim. Appl. **30**(1), 45–61 (2005)
12. Hinze, M., Volkwein, S.: Error estimates for abstract linear-quadratic optimal control problems using proper orthogonal decomposition. Comput. Optim. Appl. **39**(3), 319–345 (2007)
13. Hinze, M., Rösch, A.: Discretization of Optimal Control Problems. in G. Leugering, S. Engell, A. Griewank, M. Hinze, R. Rannacher, V. Schulz, M. Ulbrich, S. Ulbrich (eds.): Constrained Optimization and Optimal Control for Partial Differential Equations, Int. Ser. Num. Math. **160**, 391–430 (2012)
14. Kammann, E., Tröltzsch, F., Volkwein, S.: A posteriori error estimation for semilinear parabolic optimal control problems with application to model reduction by POD. ESAIM Math. Model. Num. **47**(2), 555–581 (2013)
15. Kunisch, K., Volkwein, S.: Galerkin proper orthogonal decomposition methods for parabolic problems. Numer. Math. **90**(1), 117–148 (2001)
16. Kunisch, K., Volkwein, S.: Galerkin proper orthogonal decomposition methods for a general equation in fluid dynamics. SIAM J. Numer. Anal. **40**(2), 492–515 (2002)
17. Kunisch, K., Volkwein, S.: Proper orthogonal decomposition for optimality systems. ESAIM M2AN **42**(1), 1–23 (2008)
18. Kunisch, K., Volkwein, S.: preprint available: http://nbn-resolving.de/urn:nbn:de:bsz:352-250378 Optimal snapshot location for computing POD basis functions. ESAIM M2AN **44**, 509–529 (2010)
19. Lions, J.L.: Optimal control of systems governed by partial differential equations. Grundlehren der mathematischen Wissenschaften **170**, Springer (1971)
20. Ludovici, F., Neitzel, I., Wollner, W.: A priori error estimates for state constrained semilinear parabolic optimal control problems. Submitted. INS Preprint No. 1605 (2016) available at: http://neitzel.ins.uni-bonn.de/pub/LNW2016_INS.pdf
21. Malanowski, K., Büskens, C., Maurer, H.: Convergence of approximations to nonlinear optimal control problems. in A.V. Fiacco (ed.): Mathematical programming with data perturbations. Pure Appl. Math. **195**, 253–284 (1998)
22. Meidner, D., Vexler, B.: Adaptive space-time finite element methods for parabolic optimization problems. SIAM J. Contr. Optim. **46**(1), 116–142 (2007)
23. Meidner, D., Vexler, B.: A priori error estimates for space-time finite element discretization of parabolic optimal control problems Part I: Problems without control constraints. SIAM J. Contr. Optim. **47**(3), 1150–1177 (2008)
24. Meidner, D., Vexler, B.: A priori error estimates for space-time finite element discretization of parabolic optimal control problems Part II: Problems with control constraints. SIAM J. Contr. Optim. **47**(3), 1301–1329 (2008)
25. Meidner, D., Rannacher, R., Vexler, B.: A Priori Error Estimates for Space-Time Finite Element Discretization of Parabolic Optimal Control Problems with Pointwise State Contstraints in Time. SIAM J. Contr. Optim. **49**(5), 1961–1997 (2011)

26. Meidner, D., Vexler, B.: Adaptive space-time finite element methods for parabolic optimization problems. in G. Leugering, S. Engell, A. Griewank, M. Hinze, R. Rannacher, V. Schulz, M. Ulbrich, S. Ulbrich (eds.): Constrained Optimization and Optimal Control for Partial Differential Equations, Int. Ser. Num. Math. **160**, 319–348 (2012)
27. Neitzel, I., Vexler, B.: A priori error estimates for space-time finite element discretization of semilinear parabolic optimal control problems. Numer. Math. **120**, 345–386 (2008)
28. Studinger, A., Volkwein, S.: Numerical analysis of POD a-posteriori error estimation for optimal control. in K. Bredies, C. Clason, K. Kunisch, and G. von Winckel (ed.): Control and Optimization with PDE Constraints. Int. Num. Math. **164**, 137-158 (2013)
29. Tröltzsch, F., Volkwein, S.: POD a-posteriori error estimates for linear-quadratic optimal control problems. Comp. Optim. Appl. **44**(1), 83–115 (2009)
30. Tröltzsch, F.: Optimal Control of Partial Differential Equations: Theory, Methods and Applications. Graduate Studies in Mathematics **112**, American Math. Society, Providence, Rhode Island (2010)