Model Order Reduction for PDE Constrained Optimization

Peter Benner, Ekkehard Sachs and Stefan Volkwein

Abstract. The optimization and control of systems governed by partial differential equations (PDEs) usually requires numerous evaluations of the forward problem or the optimality system. Despite the fact that many recent efforts, many of which are reported in this book, have been made to limit or reduce the number of evaluations to 5–10, this cannot be achieved in all situations and even if this is possible, these evaluations may still require a formidable computational effort. For situations where this effort is not acceptable, model order reduction can be a means to significantly reduce the required computational resources. Here, we will survey some of the most popular approaches that can be used for this purpose. In particular, we address the issues arising in the strategies discretize-then-optimize, in which the optimality system of the reduced-order model has to be solved, and optimize-then-discretize, where a reduced-order model of the optimality system has to be found. The methods discussed include versions of proper orthogonal decomposition (POD) adapted to PDE constrained optimization as well as system-theoretic methods.

Mathematics Subject Classification (2010). Primary 49M05; Secondary 93C20.

Keywords. Model order reduction, PDE constrained optimization, optimal control, proper orthogonal decomposition, adaptive methods.

1. Introduction

Optimal control problems for partial differential equations are often hard to tackle numerically because their discretization leads to very large scale optimization problems. Therefore different techniques of model reduction were

This work was supported by the DFG Priority Program 1253 “Optimization with Partial Differential Equations”, grants BE 2174/8-2, SA 289/20-1, and the DFG project ”A-posteriori-POD Error Estimators for Nonlinear Optimal Control Problems governed by Partial Differential Equations”, grant VO 1658/2-1.

developed to approximate these problems by smaller ones that are tractable with less effort.

One popular model reduction technique for large-scale state-space systems is the *moment matching approximation* considered first in [28, 30]. This method is based on projecting the dynamical system onto Krylov subspaces computed by an Arnoldi- or Lanczos process. Krylov methods prove to be efficient for large-scale sparse systems, since only matrix-vector multiplications are required. The moment matching method shows the drawbacks that stability and passivity are not necessarily preserved in the reduced-order system and that there is no global approximation error bound; see, e.g., [7, 39]. *Balanced truncation* [93] is another well studied model reduction technique for state-space systems. This method utilizes the solutions to the two Lyapunov equations, the so-called controllability and observability Gramians. The balanced truncation method is based on transforming the state-space system into a balanced form so that its controllability and observability Gramians become diagonal and equal. Moreover, the states that are difficult to reach or to observe, are truncated. The advantage of this method is that it preserves the asymptotic stability in the reduced-order system. Furthermore, a-priori error bounds are available. Recently, the theory of balanced truncation model reduction was extended to descriptor systems; see, e.g., [61] and [43]. Both the moment matching approximation and the balanced truncation approach do not rely on snapshots, which have to be taken more or less arbitrarily. For an overview we refer the reader to [3, 78]. However, up to now, both strategies can be applied more or less only to linear, time-invariant dynamical systems and do not yet cover time variant or nonlinear models. There are attempts to deal with time variant equations by approximating them through piecewise constant models; see, e.g., [14].

Recently the application of *reduced-order models* to linear time varying and nonlinear systems, in particular to nonlinear control systems, has received an increasing amount of attention. The reduced-order approach is based on projecting the dynamical system onto subspaces consisting of basis elements that contain characteristics of the expected solution. This is in contrast to, e.g., finite element techniques, where the basis elements of the subspaces do not relate to the physical properties of the system that they approximate. The *reduced basis* (RB) method, as developed in [35, 66] and [51], is one such reduced-order method, where the basis elements correspond to the dynamics of expected control regimes. Let us refer to the [26, 44, 63, 68] for the successful use of reduced basis method in PDE constrained optimization problems. Currently *Proper orthogonal decomposition* (POD) is probably the mostly used and most successful model reduction technique for nonlinear optimal control problems, where the basis functions contain information from the solutions of the dynamical system at pre-specified time-instances, so-called snapshots. Due to a possible linear dependence or almost linear dependence the snapshots themselves are not appropriate as a basis. Hence a singular value decomposition is carried out and the leading generalized eigenfunctions

are chosen as a basis, referred to as the POD basis. POD is successfully used in a variety of fields including fluid dynamics, coherent structures [1, 4] and inverse problems [8]. Moreover in [6] POD is successfully applied to compute reduced-order controllers. The relationship between POD and balancing was considered in [56, 75, 91]. An error analysis for non-linear dynamical systems in finite dimensions were carried out in [71] and a missing point estimation in models described by POD was studied in [5]. Let us also mention that POD and the reduced basis method are successfully combined by variants of the POD greedy algorithm; see [42] and [41], for instance.

Reduced order models are used in PDE-constrained optimization in various ways; see, e.g., [37, 48, 76] for a survey. In optimal control problems it is sometimes necessary to compute a feedback control law instead of a fixed optimal control. In the implementation of these feedback laws models of reduced order can play an important and very useful role, see [2, 6, 31, 55, 58, 72]. Another useful application is the use in optimization problems, where a PDE solver is part of the function evaluation. Obviously, thinking of a gradient evaluation or even a step-size rule in the optimization algorithm, an expensive function evaluation leads to an enormous amount of computing time. Here, the reduced order model can replace the system given by a PDE in the objective function. It is quite common that a PDE can be replaced by a five- or ten-dimensional system of ordinary differential equations. This results computationally in a very fast method for optimization compared to the effort for the computation of a single solution of a PDE. There is a large amount of literature in engineering applications in this regard, we mention only the papers [60, 64]. Recent applications can also be found in finance using the RB model [67] and the POD model [80] in the context of calibration for models in option pricing.

To explain the reduced-order modelling we choose the following generic nonlinear optimal control problem: Minimize the cost functional

$$J(y,u) = \frac{1}{2}\int_0^T \int_\Omega |y(t,\boldsymbol{x}) - y_d(t,\boldsymbol{x})|^2 \,\mathrm{d}\boldsymbol{x}\mathrm{d}t + \frac{\kappa}{2}\,\|u\|_U^2 \tag{1.1a}$$

subject to the semilinear parabolic partial differential equation

$$\begin{aligned} y_t(t,\boldsymbol{x}) - \Delta y(t,\boldsymbol{x}) + f(t,\boldsymbol{x},y(t,\boldsymbol{x})) &= (\mathcal{B}u)(t,\boldsymbol{x}), && (t,\boldsymbol{x}) \in Q, \\ y(t,\boldsymbol{x}) &= 0, && (t,\boldsymbol{x}) \in \Sigma, \\ y(0,\boldsymbol{x}) &= y_0(\boldsymbol{x}), && \boldsymbol{x} \in \Omega \end{aligned} \tag{1.1b}$$

and to the control constraints

$$u \in U_{ad}. \tag{1.1c}$$

In (1.1) we suppose that $T > 0$ holds and the spatial domain $\Omega \subset \mathbb{R}^d$, $d \in \{1,2,3\}$, is a bounded open set with Lipschitz-continuous boundary Γ. We set $Q = (0,T) \times \Omega$ and $\Sigma = (0,T) \times \Gamma$. Furthermore, the set $U_{ad} \subseteq U$ of admissible controls is a closed and convex subset of a Hilbert space U, which is identified by its dual space U'. Let us refer to the recent paper [38], where the authors investigate a state-constrained linear-quadratic optimal control

problem using proper orthogonal decomposition. Let us set $H = L^2(\Omega)$ and $V = H_0^1(\Omega)$ with dual space $V' = H^{-1}(\Omega)$. In (1.1a) we suppose that y_d belongs to $L^2(0,T;H)$ and $\kappa > 0$ holds. Let $f : Q \times \mathbb{R} \to \mathbb{R}$ contain the semilinear term and the control operator $\mathcal{B} : U \to L^2(0,T;H)$ be linear and continuous. Finally, we suppose that $y_0 \in L^2(\Omega)$. For given control $u \in U_{ad}$ a solution to (1.1b) is understood as a weak solution, i.e., y satisfies

$$\begin{aligned} &\frac{\mathrm{d}}{\mathrm{d}t}\,\langle y(t),\varphi\rangle_H + \int_\Omega \nabla y(t)\cdot\nabla\varphi + \big(f(t,\cdot,y(t)) - (\mathcal{B}u)(t)\big)\varphi\,\mathrm{d}\boldsymbol{x} = 0 \\ &\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \forall\varphi\in V,\ t\in(0,T], \\ &\langle y(0),\phi\rangle_H = \langle y_0,\phi\rangle_H \qquad\qquad\qquad\qquad \forall\phi\in H, \end{aligned} \tag{1.1b'}$$

where $y(t)$ stands for $y(t,\cdot)$ as a function in the spatial variable $\boldsymbol{x}$. We suppose that (1.1b') admits a unique weak solution $y = y(u)$ for any $u \in U_{\mathsf{ad}}$. We refer to [19, 73] for sufficient conditions, but also for the proof that (1.1) possesses a local optimal solution $\bar{x} = (\bar{y},\bar{u})$. Let us introduce the so called reduced cost functional[1]

$$\hat{J}(u) = J(y(u),u), \quad u \in U_{\mathsf{ad}},$$

where $y(u)$ is the unique weak solution to (1.1b').

The paper is organized as follows: First we consider two venues how to discretize a PDE-constrained optimization problem followed by an introduction to POD-based methods. We address various aspects about the proper choice of the POD basis in the course of the optimization. The fourth section is devoted to system-theoretic aspects such as techniques based on Krylov subspaces and system balancing.

2. Optimization with surrogate models

The reduced-order approximation to (1.1) can be derived by a *discretize-then-optimize* or by a *optimize-then-discretize* approach. In the first case, the optimal control problem is projected onto the reduced-order subspace and the resulting low-dimensional optimal control problem is solved by appropriate optimisation methods., like, e.g., a sequential quadratic programming algorithm [65, Chapter 18]. In the second case, we start by deriving optimality conditions for (1.1) and discretize the obtained conditions by a reduced-order projection.

Next we describe both approaches for (1.1). Suppose that $\psi_1,\ldots,\psi_\ell \in V$ are given linearly independent functions in V. Then, we define the subspace $V^\ell = \operatorname{span}\{\psi_1,\ldots,\psi_\ell\}$. Let us mention that we avoid a discretization of the control space U; see [47].

[1]Not to be confused with the "reduced-order" terminology used in the model reduction context — here, "reduced" means that the cost functional is written in dependence of the control only, using the fact that the weak solution $y(u)$ is uniquely determined by the chosen u!

2.1. Discretize-then-optimize

In this approach we only replace the state y by a reduced-order approximation. For that reason we introduce the affine Galerkin ansatz

$$y^\ell(t) = y_{\mathsf{p}}(t) + \sum_{i=1}^{\ell} \mathrm{y}_i^\ell(t)\psi_i, \quad t \in [0,T], \tag{2.1}$$

with a chosen particular element $y_{\mathsf{p}}(t) \in V$ and coefficient functions $\mathrm{y}_i^\ell : [0,T] \to \mathbb{R}$, $1 \le i \le \ell$. Then, the reduced-order Galerkin projection for (1.1) reads as follows:

$$\min J(y^\ell, u) = \int_0^T \int_\Omega \left| y^\ell(t) - y_d(t) \right|^2 \mathrm{d}\boldsymbol{x}\mathrm{d}t + \frac{\kappa}{2} \|u\|_U^2 \tag{2.2a}$$

subject to the projected nonlinear dynamical system

$$\begin{aligned} &\frac{\mathrm{d}}{\mathrm{d}t} \langle y^\ell(t), \psi \rangle_H + \int_\Omega \nabla y^\ell(t) \cdot \nabla \psi + \big(f(t, \cdot, y^\ell(t)) - (\mathcal{B}u)(t) \big) \psi \, \mathrm{d}\boldsymbol{x} \\ &\qquad\qquad \forall \psi \in V^\ell, \ t \in (0,T], \\ &\langle y^\ell(0), \psi \rangle_H = \langle y_0, \psi \rangle_H \qquad\qquad \forall \psi \in V^\ell \end{aligned} \tag{2.2b}$$

and

$$u \in U_{ad}. \tag{2.2c}$$

Notice that (2.2b) is a finite-dimensional system of ordinary differential equations for the coefficient vector $\mathrm{y}^\ell = (\mathrm{y}_1^\ell, \ldots, \mathrm{y}_\ell^\ell)^T : [0,T] \to \mathbb{R}^\ell$. Throughout this work we suppose that there exists a unique weak solution $y^\ell = y^\ell(u)$ to (2.2b) for any $u \in U_{\mathsf{ad}}$. Then, we can define the reduced cost functional associated with the reduced-order approximation as follows:

$$\hat{J}^\ell(u) = J(y^\ell(u), u), \quad u \in U_{\mathsf{ad}},$$

where $y^\ell(u)$ denotes the weak solution to (2.2b).

2.2. Optimize-then-discretize

Using a Lagrangian framework it is straightforward to derive first-order necessary optimality conditions for (1.1); see, e.g., [46, Chapter 1]. Assuming that $(\bar{y}, \bar{u})$ is a local optimal solution to (1.1) and that a constraint qualification condition is satisfied there exists a *Lagrange multiplier* or *dual variable* $\bar{p}$ satisfying together with $\bar{x} = (\bar{y}, \bar{u})$ the following adjoint or dual equation [86, Chapter 6]

$$\begin{aligned} &-\frac{\mathrm{d}}{\mathrm{d}t} \langle \bar{p}(t), \varphi \rangle_H + \int_\Omega \nabla \bar{p}(t) \cdot \nabla \varphi + f_y(t, \cdot, \bar{y}(t)) \bar{p}(t) \varphi \, \mathrm{d}\boldsymbol{x} \\ &\qquad = \int_\Omega \big(y_d(t) - \bar{y}(t) \big) \varphi \, \mathrm{d}\boldsymbol{x} \qquad \forall \varphi \in V, \ t \in (0,T], \\ &\langle \bar{p}(T), \phi \rangle_H = 0 \qquad\qquad \forall \phi \in H \end{aligned} \tag{2.3a}$$

and the *variational inequality*

$$\langle \kappa \bar{u} - \mathcal{B}^\star \bar{p}, u - \bar{u} \rangle_U \ge 0 \quad \forall u \in U_{ad} \tag{2.3b}$$

where $\mathcal{B}^\star : L^2(0,T;H) \to U$ denotes the adjoint operator of $\mathcal{B}$. Now, the state equation (1.1b') as well as the system (2.3) are discretized by a reduced-order Galerkin scheme. Here, we choose the same Galerkin ansatz functions as for the state variable which its motivated by the error analysis in [37]. Analogous to (2.1) we make the ansatz

$$p^\ell(t) = p_{\mathsf{p}}(t) + \sum_{i=1}^{\ell} \mathrm{p}_i^\ell(t)\psi_i, \quad t \in [0,T], \tag{2.4}$$

for the adjoint variable, where $p_{\mathsf{p}}(t) \in V$ is a chosen particular function and $\mathrm{p}_i^\ell : [0,T] \to \mathbb{R}$ stands for the ℓ modal coefficient functions. Then, we arrive at the following reduced-order system for the unknown reduced-order solution $\bar{x}^\ell = (\bar{y}^\ell, \bar{u}^\ell)$ and $\bar{p}^\ell$

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}t}\langle \bar{y}^\ell(t), \psi\rangle_H + \int_\Omega \nabla \bar{y}^\ell(t)\cdot\nabla\psi + \big(f(t,\cdot,\bar{y}^\ell(t)) - (\mathcal{B}\bar{u}^\ell)(t)\big)\psi \,\mathrm{d}\boldsymbol{x},\\
&\langle y^\ell(0),\psi\rangle_H = \langle y_0,\psi\rangle_H,\\
-&\frac{\mathrm{d}}{\mathrm{d}t}\langle \bar{p}^\ell(t), \psi\rangle_H + \int_\Omega \nabla \bar{p}^\ell(t)\cdot\nabla\psi + \big(f_y(t,\cdot,\bar{y}^\ell(t))\bar{p}^\ell(t) + \bar{y}(t) - y_d(t)\big)\psi \,\mathrm{d}\boldsymbol{x},\\
&\langle \bar{p}^\ell(T),\psi\rangle_H = 0
\end{aligned}$$

for all $\psi \in V^\ell$, $t \in [0,T]$ and

$$\langle \kappa\bar{u}^\ell - \mathcal{B}^\star\bar{p}^\ell, u - \bar{u}^\ell\rangle_U \geq 0 \quad \forall u \in U_{ad}.$$

To solve the obtained reduced-order scheme for the first-order necessary optimality conditions one can apply, e.g., semismooth Newton [45] or interior point methods [77, 88].

3. POD-Based Methods

Let X be either the space H or the space V. In X we denote by $\langle\cdot,\cdot\rangle_X$ and $\|\cdot\|_X = \langle\cdot,\cdot\rangle_X^{1/2}$ the inner product and the associated norm, respectively. Notice that X is separable, i.e., X has a countable dense subset. This implies that X possesses a countable orthonormal basis; see, e.g., [74, p. 47].

For fixed $n, \wp \in \mathbb{N}$ let the so-called *snapshots* $w_1^k, \ldots, w_n^k \in X$ be given for $1 \leq k \leq \wp$. To avoid a trivial case we suppose that at least one of the w_j^k's is nonzero. Then, we introduce the finite-dimensional, linear subspace

$$\mathcal{V} = \operatorname{span}\Big\{ w_j^k \,|\, 1 \leq j \leq n \text{ and } 1 \leq k \leq \wp \Big\} \subset X \tag{3.1}$$

with dimension $d \in \{1, \ldots, n\wp\}$. We call the set $\mathcal{V}$ *snapshot subspace*. The method of POD consists in choosing a complete orthonormal basis $\{\psi_i\}_{i=1}^\infty$ in X such that for every $\ell \in \{1,\ldots,d^n\}$ the mean square error between the $n\wp$ elements y_j^k and their corresponding ℓ-th partial Fourier sum is minimized

on average:

$$\begin{cases} \min \sum_{k=1}^{\wp} \sum_{j=1}^{n} \alpha_j \left\| w_j^k - \sum_{i=1}^{\ell} \langle w_j^k, \psi_i \rangle_X \psi_i \right\|_X^2 \\ \text{s.t. } \{\psi_i\}_{i=1}^{\ell} \subset X \text{ and } \langle \psi_i, \psi_j \rangle_X = \delta_{ij},\ 1 \le i, j \le \ell, \end{cases} \tag{3.2}$$

where the α_j's denote positive weighting parameters. Here, the symbol δ_{ij} denotes the Kronecker symbol satisfying $\delta_{ii} = 1$ and $\delta_{ij} = 0$ for $i \neq j$. An optimal solution $\{\bar{\psi}_i^n\}_{i=1}^{\ell}$ to (3.2) is called a *POD basis of rank* ℓ.

To solve (3.2) we define the linear operator $\mathcal{R} : X \to X$ as follows:

$$\mathcal{R}\psi = \sum_{k=1}^{\wp} \sum_{j=1}^{n} \alpha_j \langle \psi, w_j^k \rangle_X w_j^k \quad \text{for } \psi \in X \tag{3.3}$$

with positive weights $\alpha_1, \ldots, \alpha_n$. Then, $\mathcal{R}$ is a compact, nonnegative and self-adjoint operator. Suppose that $\{\bar{\lambda}_i\}_{i=1}^{\infty}$ and $\{\bar{\psi}_i^n\}_{i=1}^{\infty}$ denote the nonnegative eigenvalues and associated orthonormal eigenfunctions of $\mathcal{R}$ satisfying

$$\mathcal{R}\bar{\psi}_i = \bar{\lambda}_i \bar{\psi}_i, \quad \bar{\lambda}_1 \ge \ldots \ge \bar{\lambda}_d > \bar{\lambda}_{d+1} = \ldots = 0.$$

Then, for every $\ell \in \{1, \ldots, d\}$ the first ℓ eigenfunctions $\{\bar{\psi}_i\}_{i=1}^{\ell}$ solve (3.2). Moreover, the value of the cost evaluated at the optimal solution $\{\bar{\psi}_i\}_{i=1}^{\ell}$ satisfies

$$\sum_{k=1}^{\wp} \sum_{j=1}^{n} \alpha_j \left\| w_j^k - \sum_{i=1}^{\ell} \langle w_j^k, \bar{\psi}_i \rangle_X \bar{\psi}_i \right\|_X^2 = \sum_{i=\ell+1}^{d} \bar{\lambda}_i.$$

For more details we refer the reader to [48, 50] and [37, Chapter 2], for instance.

Remark 3.1. a) In the context of the optimal control problem (1.1) a reasonable choice for the snapshots is $w_j^1 \approx y(t_j)$ and $w_j^2 \approx p(t_j)$ for time grid $0 = t_1 < \ldots < t_n = T$. Utilizing new POD error estimates for evolution problems [20, 82] and optimal control problems [49, 87] convergence and rate of convergence results are derived for linear-quadratic control constrained problems in [37] for the choices $X = H$ and $X = V$.

b) For the numerical realization the space X has to be discretized by, e.g., finite element discretizations. In this case the Hilbert space X has to be replaced by an Euclidean space $\mathbb{R}^m$ endowed with a weighted inner product; see [37].

3.1. A-Posteriori Error Analysis

In contrast to methods of balanced truncation type, the POD method is somehow lacking a reliable a-priori error analysis. Unless its snapshots are generating a sufficiently rich state space, it is not a-priorly clear how far the optimal solution of the POD problem is from the exact one. On the other hand, the POD method is a universal tool that is applicable also to problems with time-dependent coefficients or to nonlinear equations. Moreover, by generating snapshots from the real (large) model, a space is constructed that inhibits the main and relevant physical properties of the state system. This,

and its ease of use makes POD very competitive in practical use, despite of a certain heuristic flavor.

Based on a perturbation argument [27] it is derived in [53, 87] how far the suboptimal control $\bar{u}^\ell$, computed on the basis of the POD model, is from the (unknown) exact $\bar{u}$. Let $\mathcal{D}$ be an open, bounded subset of $\mathbb{R}^{\mathsf{p}}$ and $U = L^2(\mathcal{D})$. By $U_{\mathsf{ad}} \subset U$ we define the closed, convex and bounded subset

$$U_{\mathsf{ad}} = \big\{ u \in L^2(\mathcal{D}) \,\big|\, u_a(s) \le u(s) \le u_b(s) \text{ for almost all } s \in \mathcal{D} \big\}$$

with $u_a, u_b \in L^2(\mathcal{D})$ satisfying $u_a \le u_b$ in $\mathcal{D}$ a.e. Suppose that $f \equiv 0$ holds, i.e., (1.1b) is a linear evolution problem. Then, the error estimate reads as follows:

$$\|\bar{u}^\ell - \bar{u}\|_U \le \frac{1}{\kappa} \|\zeta^\ell\|_U \tag{3.4}$$

where the computable perturbation function $\zeta^\ell \in U$ is given by

$$\zeta^\ell = \begin{cases} -\min\big(0, \kappa\bar{u}^\ell - \mathcal{B}^\star\tilde{p}^\ell\big) & \text{in } \mathcal{A}_a^\ell = \big\{ s \in \mathcal{D} \,\big|\, \bar{u}^\ell = u_a \text{ a.e.} \big\}, \\ \max\big(0, \kappa\bar{u}^\ell - \mathcal{B}^\star\tilde{p}^\ell\big) & \text{in } \mathcal{A}_b^\ell = \big\{ s \in \mathcal{D} \,\big|\, \bar{u}^\ell = u_b \text{ a.e.} \big\}, \\ -\big(\kappa\bar{u}^\ell - \mathcal{B}^\star\tilde{p}^\ell\big) & \text{in } \mathcal{D} \setminus \big(\mathcal{A}_a^\ell \cup \mathcal{A}_b^\ell\big). \end{cases}$$

Futhermore, $\tilde{y}$ and $\tilde{p}$ solve

$$\begin{aligned} &\frac{\mathrm{d}}{\mathrm{d}t} \langle \tilde{y}^\ell(t), \varphi \rangle_H + \int_\Omega \nabla\tilde{y}^\ell(t) \cdot \nabla\varphi - (\mathcal{B}\bar{u}^\ell)(t)\big)\varphi \,\mathrm{d}\boldsymbol{x}, \\ &\langle \tilde{y}^\ell(0), \psi \rangle_H = \langle y_0, \psi \rangle_H, \\ -&\frac{\mathrm{d}}{\mathrm{d}t} \langle \tilde{p}^\ell(t), \varphi \rangle_H + \int_\Omega \nabla\tilde{p}^\ell(t) \cdot \nabla\varphi + \big(\tilde{y}^\ell(t) - y_d(t)\big)\varphi \,\mathrm{d}\boldsymbol{x}, \\ &\langle \tilde{p}^\ell(T), \varphi \rangle_H = 0 \end{aligned}$$

for all $\varphi \in V$, $t \in [0, T]$. It is shown in [37, 87] that $\|\zeta^\ell\|_U$ tends to zero as ℓ tends to infinity. Hence, increasing the number of POD ansatz functions leads to more accurate POD suboptimal controls. This idea turns out to be numerically very efficient. For linear-quadratic problems we refer to [37, 38, 84, 85, 90]. It is able to compensate for the lack of a priori analysis for POD methods. The analysis is extended to nonlinear optimal control problems in [26, 53]. For a parameter estimation example we also refer to [57]. Unfortunately, the a-posteriori error estimates requires a lower bound for the smallest eigenvalue of the reduced Hessian, which is – unless the control space is low dimensional – usually computationally expensive. Another approach is to use the a-posteriori error estimate in an inexact sequential quadratic programming (SQP) approach, where the sufficient accuracy of each level of the SQP method is guaranteed by the error control; see [52]. Let us mention that there is related work also available in the reduced-basis literature; see, e.g., [24, 34, 63]. Here, the authors concentrate in deriving online efficient a-posteriori error estimators.

3.2. Optimality System POD

The accuracy of the reduced-order model can be controlled by the a-posteriori error analysis presented in the previous subsection. However, if the POD basis is created from a reference trajectory containing features which are quite different from those of the optimally controlled trajectory, a rather huge number of POD ansatz functions have to be included in the reduced-order model. This fact may lead to non-efficient reduced-order models and numerical instabilities. To avoid these problems the POD basis is generated in an initialization step utilizing *optimality system POD* (OS-POD); see [54]. In OS-POD the POD basis is updated in the direction of the minimum of the cost. Recall that the POD basis is computed from the state $y = y(u)$ with some control $u \in \mathcal{U}_{\mathsf{ad}}$. Thus, the reduced-order Galerkin projection depends on the state variable and hence on the control u at which the eigenvalue $\mathcal{R}\psi_i = \lambda_i \psi_i$ for $i = 1, \ldots, \ell$ is solved for the basis $\{\psi_i\}_{i=1}^{\ell}$. This may deter from one of the main advantages of the POD approach for model reduction, which consists in the fact that unlike typical finite element basis functions the elements of the POD basis reflects the dynamics of the system. In optimal control this feature gets lost if the dynamics of the state corresponding to the reference control is significantly different from the trajectory corresponding to the optimal approach. Hence, we propose to consider the extended problem [54]:

$$\min J(y^\ell, u) \text{ s.t. } \begin{cases} z = (y^\ell, y, \lambda_i, \psi_i), u \in U_{ad} \\ (y^\ell, u) \in H^1(0,T; V^\ell) \times U_{\mathsf{ad}} \text{ satisfy (2.2b)}, \\ (y, u) \text{ satisfy (1.1b')}, \\ \mathcal{R}(y)\psi_i = \lambda_i \psi_i, \ 1 \le i \le \ell. \end{cases} \qquad (\mathbf{P}^\ell_{\text{ospod}})$$

Notice that the second line of the constraints in $(\mathbf{P}^\ell_{\text{ospod}})$ coincide with the constraints in (2.2), the next two are the infinite-dimensional state equation and the eigenvalue problem characterizing the POD basis. For the optimal solution the problem formulation $(\mathbf{P}^\ell_{\text{ospod}})$ has the property that the associated POD reduced system is computed from the trajectory corresponding to the optimal control and thus, differently from (2.2), the problem of unmodelled dynamics is removed. Of course, $(\mathbf{P}^\ell_{\text{ospod}})$ is more complicated than (2.2). For practical realization an operator splitting approach is used in [54], where also sufficient conditions are given so that $(\mathbf{P}^\ell_{\text{ospod}})$ possesses a unique optimal solution $(\bar{y}^\ell, \bar{y}, \bar{\lambda}_i, \bar{\psi}_i, \bar{u}^\ell)$, which can be characterized by first-order necessary optimality conditions. Convergence results for OSPOD are studied in the Ph.D thesis [62]. The combination of OS-POD and a-posteriori error analysis is investigated in the paper [89] and the recent master thesis [36].

3.3. Trust Region POD

In PDE-constrained optimization the use of reduced order models is highly efficient, since the complex and time-consuming constraint in form of a partial

differential equation is replaced by a relatively small system of ordinary differential equations. There is, however, one caveat that needs to be addressed. Suppose, one uses a reduced order model as in (2.2a)-(2.2b), where the state y_k^ℓ is based on a reduced order model at a certain control vector u_k. Then it is usually numerically very efficient to minimize this reduced order model and obtain a solution $(\bar{y}^\ell, \bar{u})$.

$$J(\bar{y}, \bar{u}) \leq J(y_k^\ell, u) \text{ for all } (y_k^\ell, u) \text{ that satisfy (2.2b).}$$

The problem that might occur is the fact that during this minimization one usually moves away from the original control u_k on which the reduced order model was built. In this case, it could happen that the quality of the model deteriorates and, in particular, $\bar{y}^\ell$ is no longer a good approximation of the solution $\bar{y}$ to the full model.

In [1], a direct estimate is used to monitor the accuracy of the model. In [4] a trust region approach was proposed to manage updating the reduced order model. The trust region method is a well known method for the globalization of locally convergent methods like Newton's method. Here the quadratic Taylor expansion is used as a model function for the nonlinear objective function. However, it is known that this globalization strategy also works for nonlinear models. Hence, one could use it also in this context, where a nonlinear model is replaced by another nonlinear model, e.g. POD, but of much smaller complexity than the original one.

The main key for controlling the quality of the model approximation in the trust region framework is the comparison between the predicted reduction starting from full model evaluation at (y_k, u_k)

$$v_{pred} = \hat{J}^\ell(\bar{u}) - \hat{J}(u_k)$$

and the actual reduction

$$v_{actual} = \hat{J}(\bar{u}) - \hat{J}(u_k),$$

where $\bar{y}$ is the true solution at the control $\bar{u}$. If the quotient $\rho = v_{pred}/v_{actual}$ is close to 1, then the reduced model is a good model and we accept $\bar{u}$ as u_{k+1} and likewise $\bar{y}$ as y_{k+1}. If ρ differs significantly from 1, we have to modify the model, for example, by introducing more snapshots or restricting the controls to a ball around u_k.

One drawback of this strategy is the evaluation of the full model at the control u_*. This problem can be alleviated by the introduction of a hierarchy of models, e.g. created by different fine discretizations of the PDE, and the full model in this case is replaced by the next finer model. In this way, when approaching the finest level in the hierarchy, we already are very close to the optimal point. This approach is also closely related to multifidelity optimization.

With respect to convergence in an optimization framework, the accuracy of the gradient of the model function has to be monitored. As it is shown in [80], a-priori-estimates on the accuracy of the POD approximation can be

used to establish error bounds for the distance between the gradient of the model function and the original function.

Trust region POD (TRPOD) is meanwhile a well established method in applications: In the control of fluid flow problems applications of TRPOD go back to Bergmann and Cordier [17]. More recently, Navon and coworkers [21] use this methodology in order to accomplish a 4-D VAR simulation for the shallow water equation. In [92], the TRPOD framework is combined with an efficient error bound for defining the trust region in the design optimization of vibrating structures using frequency domain formulations. An application of TRPOD to the optimization of simulated moving bed chromatography (a separation process in chemical process engineering) is discussed in [59], Also, in financial applications, reduced order models, see [22] or [79] are gaining importance and TRPOD is a promising tool in the calibration process.

In [33] the Carter condition

$$\|\nabla \hat{J}(u_k) - \nabla \hat{J}^\ell(u_k)\|_U \leq \sigma \|\nabla \hat{J}^\ell(u_k)\|_U, \quad \sigma \in (0,1),$$

which is essential for the convergence of the inexact trust region method, is replaced by an a posteriori error estimation in order to control the reduced-order approximation of the reduced cost functional. For that purpose error estimates for the state and the dual variable are utilized. This offers a bridge between trust region POD and the a posteriori analysis presented in Section 3.1.

4. System-Theoretic MOR Approaches

A disadvantage of POD-based methods is that the control function $u(t)$ has to be chosen in advance in order to compute the training set. As this reference control may differ significantly from the optimal control, the reduced order model may not be suitable for optimization, or the optimal control computed using the reduced order model may be too far away from the one obtained from the full-order model. The alternative OS-POD avoiding this problem is discussed in subsection 3.2. As already noted there, solving the extended problem ($\mathbf{P}^\ell_{\text{ospod}}$) is computationally challenging. The advantage of using system-theoretic approaches is that these are so-called "simulation-free" methods which do not require training sets. The quality of the reduced order model thus is independent of the chosen control function u(t). For the purpose of deriving error bounds, it is usually assumed that $U_{ad} = L^2(0,\infty;\mathbb{R}^m)$, that is, we assume m scalar input functions that are square-integrable for an infinite time horizon. It should be stressed, though, that this assumption is only required for the derivation of error bounds for the reduced order models, while the model itself is applicable of course also in the presence of control constraints or other spaces of admissible control functions.

The starting point for system-theoretic approaches to model order reduction is the description of a mathematical model as a linear or nonlinear system. As these methods were developed mainly for linear problems, and

extensions to nonlinear problems exist, but are either based on heuristics or are computationally intractable, we only discuss the application to linear systems here. But significant progress in nonlinear model reduction using system-theoretic concepts can be expected in the near future and this may then lead to useful methods also in the context of PDE constrained optimization.

Thus, in the following we will consider *linear, time-invariant (LTI) systems*

$$\begin{aligned} \mathrm{M}\dot{\mathrm{y}}(t) &= -\mathrm{S}\mathrm{y}(t) + \tilde{\mathrm{B}}\mathrm{u}(t) && \text{(4.1a)} \\ \mathrm{z}(t) &= \mathrm{C}\mathrm{y}(t), && \text{(4.1b)} \end{aligned}$$

for a given initial condition $\mathrm{y}(0) = \mathrm{y}_0 \in \mathbb{R}^N$. Here, y is the state of the system and in this context is usually the discrete vector obtained by discretization of (1.1b'), i.e., $\mathrm{y}_j(t)$ are the coefficients in the ansatz $y(t) \approx \sum_{j=1}^N \mathrm{y}_j(t)\psi_j$, with basis functions ψ_j of the chosen trial space, when evaluating $\langle y(t), \phi_j \rangle_H$ for the basis functions $\phi_j \in V_N$, $j = 1, \ldots, N$, where $V_N \subset V$ is the chosen N-dimensional space of test functions. In this setting, M and S contain the corresponding mass and stiffness matrices and $\tilde{\mathrm{B}}$ is the discretized version of the input operator $\mathcal{B}$. The second equation in (4.1) is called the *output equation*, and $\mathrm{z}(t) \in \mathbb{R}^q$ is a vector of quantities of interest which in the setting considered here might be derived from a linear cost functional $z(t) = \mathcal{L}(y)(t)$ which might be goal of optimization. In practice, it is often a model for the possible measurements of the system, i.e., it is assumed that the full state y is not accessible for measurements. In case one assumes all state information is available, one might simply set $\mathrm{C} = I_N$, the identity matrix on $\mathbb{R}^N$. We will discuss the use of the output equation in the context of PDE constrained optimization below.

As the mass matrix M is invertible, for the clarity of presentation we work with the transformed system

$$\begin{aligned} \dot{\mathrm{y}}(t) &= \mathrm{A}\mathrm{y}(t) + \mathrm{B}\mathrm{u}(t) && \text{(4.2a)} \\ \mathrm{z}(t) &= \mathrm{C}\mathrm{y}(t), && \text{(4.2b)} \end{aligned}$$

where $\mathrm{A} := -\mathrm{M}^{-1}\mathrm{S}$ and $\mathrm{B} := \mathrm{M}^{-1}\tilde{\mathrm{B}}$. It should be noted, though, that we do this only formally, that is, in all computations of reduced order models, one never forms A, B explicitly but works with the original $\tilde{\mathrm{B}}$ and the usually sparse matrices M, S in order to avoid fill-in as well as possible round-off errors implied by ill-conditioning of M.

The core observation of system-theoretic approaches is that the relation from inputs u to the outputs z can be represented algebraically using the Laplace transform, yielding

$$\mathrm{z}(s) = \mathrm{C}(sI_N - \mathrm{A})^{-1}\mathrm{B}\mathrm{u}(s) =: \mathrm{G}(s)\mathrm{u}(s), \tag{4.3}$$

where by abuse of notation, we denote the Laplace transforms of z, u again by z, u, and $s \in \mathbb{C}$ is the Laplace variable. The *transfer function matrix (TFM)* G is a complex $q \times m$ matrix for any s whose entries are scalar real rational functions of a complex variable and with degree less than or equal

to N. The particular case of a single input, single output (SISO) system with $m = q = 1$ is often of particular interest in PDE constrained optimization as many problems can be formulated such that a single quantity is to be minimized using only a scalar control function $\mathrm{u}(t)$.

Equation (4.3) shows that the output of the system can be well approximated when we can find a reduced order model with TFM G^ℓ for which $\|\mathrm{G}(.)-\mathrm{G}^\ell(.)\|$ is small in an appropriate norm. It is convenient here to use L_2-related norms as they result in L_2- or L_∞-norm error bounds for $\mathrm{z}(.)-\mathrm{z}^\ell(.)$, where z^ℓ is the output of the reduced order model. Therefore it is the aim here to find a reduced order LTI system

$$\begin{aligned} \dot{\mathrm{y}}^\ell(t) &= \mathrm{A}^\ell \mathrm{y}^\ell(t) + \mathrm{B}^\ell \mathrm{u}(t) && \text{(4.4a)} \\ \mathrm{z}^\ell(t) &= \mathrm{C}^\ell \mathrm{y}^\ell(t), && \text{(4.4b)} \end{aligned}$$

where $\mathrm{y}^\ell(t) \in \mathbb{R}^\ell$ with $\ell \ll N$ and corresponding matrices of compatible sizes such that the TFM

$$\mathrm{G}^\ell(s) = \mathrm{C}^\ell(sI_\ell - \mathrm{A}^\ell)^{-1}\mathrm{B}^\ell, \tag{4.5}$$

of (4.4) approximates G well. The strength of system-theoretic model reduction methods now shows in the fact that the reduced order model (4.4) uses the same control function $\mathrm{u}(t)$ as the full-order model (4.2). Thus, neither a discretization of the input space U nor an a priori choice of a reference control u are necessary, and the optimal control $\bar{\mathrm{u}}$ obtained from using (4.4) as surrogate in the optimization process can be directly applied in the full-order model (4.2) or even in the original PDE problem when the problem is formulated such that only the amplitude of the control signal is subject to optimization.

In the following, we will discuss the two main concepts used in model order reduction of LTI systems of the form (4.2). Both are based on Petrov-Galerkin projection, i.e., the reduced order model is computed using two full-rank matrices $V, W \in \mathbb{R}^{N\times\ell}$:

$$\mathrm{A}^\ell := W^T\mathrm{A}V, \quad \mathrm{B}^\ell := W^T\mathrm{B}, \quad \mathrm{C}^\ell := \mathrm{C}V, \tag{4.6}$$

corresponding to projecting the state-space onto range(W) and using the ansatz $\mathrm{y}(t) \approx V\mathrm{y}^\ell(t)$. The basis matrices V and W for the trial and test spaces are chosen to be bi-orthonormal, i.e., $W^TV = I_\ell$ such that VW^T becomes an oblique projector. Inserting the ansatz into (4.4) leads to a nonzero residual $V\dot{\mathrm{y}}^\ell - AV\mathrm{y}^\ell - B\mathrm{u}$ which obviously is orthogonal to range(W) as

$$W^T(V\dot{\mathrm{y}}^\ell - AV\mathrm{y}^\ell - B\mathrm{u}) = \dot{\mathrm{y}}^\ell - A^\ell\mathrm{y}^\ell - B^\ell\mathrm{u} = 0.$$

In this sense, all the considered methods in this section are Petrov-Galerkin methods and become Galerkin projection methods in case $W = V$. The two model reduction techniques discussed in the following subsections only differ in the way V, W are computed, and in the resulting theoretical properties of the reduced order model. For more details on system-theoretic model order reduction techniques, consult, e.g., the recent monographs, edited volumes

and survey papers [3, 10, 14, 78]. Also note that extensions to so-called *descriptor systems*, where M in (4.1) is singular, are possible, see, e.g., [14, Chapter 3].

4.1. Rational Interpolation Based Techniques

The first family of system-theoretic model order reduction methods is based on (rational) interpolation of the TFM. The interpolant is chosen as a rational matrix function of lower degree satisfying certain interpolation conditions. Hence, the original and reduced order TFMs (and some of their first derivatives) coincide:

$$\frac{d^j}{ds^j}\mathrm{G}(s_k) = \frac{d^j}{ds^j}\mathrm{G}^\ell(s_k), \quad k = 0, \ldots, K, \quad j = 0, \ldots, J_k, \tag{4.7}$$

where the interpolation points s_k are chosen such that $(\mathrm{A} - s_k I_N)$ is nonsingular. Practically this is usually realized by certain Krylov subspace methods.

The classical approach using $K = 0$ and a sufficiently large J_K, leading to rational Hermite interpolation at s_0, has become popular as *moment matching* or *Padé(-type) approximation* since the mid-1990ies. These methods can be derived by power series expansions of the TFMs of the original and reduced order systems about s_0. The reduced order model is then determined so that the first coefficients in the series expansions match. In this context, the coefficients of the power series expansions are called moments, explaining the name "moment matching". Padé-approximation in this context means that the number of matching moments is maximized for a given degree of the approximating rational function G^ℓ. The observation that a reduced order model with the moment matching property is obtained by applying r steps of the (block) Arnoldi or Lanczos processes to $(\mathrm{A} - s_0 I_N)^{-1}$ or its (real) transpose with B or C^T as starting (block) vector and using bi-orthonormal bases of the resulting Krylov subspaces $\mathcal{K}_r((\mathrm{A} - s_0 I_N)^{-1}, \mathrm{B})$ and $\mathcal{K}_r((\mathrm{A}^T - s_0 I_N)^{-1}, \mathrm{C}^T)$ for V and W was the breakthrough of this approach as model reduction method — previous attempts employing explicit computations of the moments were so prone to round-off errors that they could not be used in practice, see [28, 30] for details.

For the SISO case, one obtains $r = \ell$ and $J_0 = 2r-1$. The bi-orthonormal bases for the two Krylov subspaces are obtained automatically using the two-sided (unsymmetric) Lanczos process. One can also run the standard Arnoldi process for both Krylov subspaces independently and enforce bi-orthonormality afterwards to obtain the same results. Using only one of the Krylov subspaces and $V = W$, one has only $J_0 = r - 1$, but the computation can be performed with the stable Arnoldi process and orthogonal projection can be used to compute the reduced order model. This is beneficiary if A is, e.g., negative definite as for stiffness matrices resulting from the Galerkin finite element methods for linear diffusion-reaction problems, as in this case, negative definiteness of A^ℓ is guaranteed.

For $m, q > 1$, things become a lot more complicated, starting from the fact that V, W necessarily need to be of the same size, but the block Krylov

subspaces will usually differ in dimension whenever $m \neq q$ (and often even if $m = q$ due to deflation). Computations in this case can be performed using block or band Lanczos/Arnoldi processes. For many more details on this, see the monographs and surveys already mentioned above or [29]

Also observe that the use of an expansion point $s_0 \notin \mathbb{R}$ will lead to a complex-valued reduced order system (4.4). Often, this is undesired. An easy remedy is to use two expansions point $s_0, s_1 = \overline{s_0}$ and to concatenate the resulting bases. If performed with care, the resulting V, W are then real.

Note that interpolation at $s_0 = \infty$ is also possible. In that case, one computes Krylov subspaces for (A, B) and/or $(\mathrm{A}^T, \mathrm{C}^T)$. Then the moments are called *Markov parameters* and the "moment matching" problem is known as *partial realization*.

The use of a single expansion point s_0 leads to good approximation only locally. Hence, it is often useful to use more than one expansion point, yielding multi-point moment matching methods, also called *rational Krylov methods*, see, e.g., [3, 10]. By fixing $K = \ell - 1$ in (4.7), in [40] a fix-point iteration to determine locally optimal expansion points w.r.t. approximation of the TFM in the H_2-norm (basically, this is the L_2-norm of the TFM evaluated on the whole imaginary axis) is suggested. In the SISO case, if this iteration converges, the obtained rational Krylov subspaces yield basis matrices V, W such the reduced order model satisfies (4.7) for $J_k = 1$ for all $k = 0, \ldots, K$ at the mirror images (w.r.t. the imaginary axis) of the eigenvalues of A^ℓ, i.e., the poles of the reduced order TFM. This property of a reduced TFM is known to be the necessary condition for a local minimizer of $\|\mathrm{G} - \mathrm{G}^\ell\|_{H_2}$.

Some extensions of the presented approaches to nonlinear systems are discussed in the recent survey [10].

4.2. Methods Based on System Balancing

Balanced Truncation (BT) has been the workhorse for model reduction in systems and control theory, in particular for controller design, for the last three decades, see, e.g., [3, 10] for a thorough discussion and references to original work. Its advantages over most other methods is that for asymptotically stable models it guarantees stability of the reduced-order model — a property that none of the other methods discussed here possesses unless additional assumptions are posed or a post-processing step is included — and it has a computable a priori error bound that allows the adaptive selection of the reduced model order given a user-defined error threshold. The method is based on two main ingredients: balancing and truncation. In the following, we will briefly introduce these concepts, starting with the latter one.

The concept of *truncation* is based on finding a suitable state-space transformation $\mathcal{T}$ defined by a nonsingular matrix $\mathrm{T} \in \mathbb{R}^{n \times n}$:

$$\begin{aligned} \mathrm{T}^{-1}\dot{\mathrm{y}}(t) &= (\mathrm{T}^{-1}\mathrm{A}\mathrm{T})\mathrm{T}^{-1}\mathrm{y}(t) + (\mathrm{T}^{-1}\mathrm{B})\mathrm{u}(t) && \text{(4.8a)} \\ \mathrm{z}(t) &= (\mathrm{C}\mathrm{T})\mathrm{T}^{-1}\mathrm{y}(t) && \text{(4.8b)} \end{aligned}$$

with partitioning

$$\mathrm{T}^{-1}\mathrm{A}\mathrm{T} = \begin{bmatrix} \mathrm{A}_{11} & \mathrm{A}_{12} \\ \mathrm{A}_{21} & \mathrm{A}_{22} \end{bmatrix}, \quad \mathrm{T}^{-1}\mathrm{B} = \begin{bmatrix} \mathrm{B}_1 \\ \mathrm{B}_2 \end{bmatrix}, \quad \mathrm{C}\mathrm{T} = [\mathrm{C}_1 \quad \mathrm{C}_2], \tag{4.9}$$

where $A_{11} \in \mathbb{R}^{\ell \times \ell}$ and the other blocks are of compatible sizes. The reduced-order model is then simply obtained by truncating the states $\ell + 1, \ldots, N$ of the transformed state vector $\mathrm{T}^{-1}\mathrm{y}(t)$ i.e., setting them to zero so that

$$\mathrm{A}^\ell = \mathrm{A}_{11}, \quad \mathrm{B}^\ell = \mathrm{B}_1, \quad \mathrm{C}^\ell = \mathrm{C}_1. \tag{4.10}$$

The challenge, of course, is to find a $\mathcal{T}$ yielding a reduced-order model with good approximation properties. Here is where the second ingredient, *balancing*, comes into the game.

For the following, we assume that A is asymptotically stable, i.e., all its eigenvalues are in the left half plane. This implies that the transfer function $\mathrm{G}(s)$ of (4.2) has all its poles in the open left half plane, thus such systems are called (asymptotically) stable. Extensions to unstable systems are possible, but will not be discussed here for brevity (see [11] for some more details and references and [13] for a recent application to the control of unstable flow problems). Strictly speaking, a balancing transformation $\mathcal{T}_b$ yields a realization of an asymptotically stable LTI system (4.2) in the form (4.8) such the unique solutions $\mathrm{P}, \mathrm{Q} \in \mathbb{R}^{N \times N}$ of the *Lyapunov equations*

$$\mathrm{A}\mathrm{P} + \mathrm{P}\mathrm{A}^T + \mathrm{B}\mathrm{B}^T = 0, \quad \mathrm{A}^T\mathrm{Q} + \mathrm{Q}\mathrm{A} + \mathrm{C}^T\mathrm{C} = 0 \tag{4.11}$$

are equal and diagonal, $\mathrm{P} = \mathrm{Q} = \mathrm{diag}\{\sigma_1, \ldots, \sigma_N\}$, with decaying *Hankel singular values* σ_k, i.e., $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_N > 0$. Such a balancing transformation does not always exist, it requires some additional system-theoretic assumption (see, e.g., [3] for a full account of these), but even if does not exist, it is possible to obtain a partial transformation that suffices to compute a reduced-order model with exactly the same properties as described in the following, see [15]. If one now assumes that a balancing transformation was used to compute (4.9) and the reduced-order model is then obtained via truncation as in (4.10), it holds:

1. A^ℓ and thus G^ℓ are asymptotically stable, and
2. the reduced-order model satisfies the error bound

$$\|\mathrm{z} - \mathrm{z}^\ell\|_{L_2} \leq \sum_{k=\ell+1}^{n} \sigma_j \|\mathrm{u}\|_{L_2}.$$

As the Hankel singular values can be computed as a by-product of the balancing transformation, it is possible to use this error bound to adapt the size of the reduced-order model to match a desired maximal approximation error.

It should be noted that in practical computations, neither the full transformation matrix T nor the solutions P, Q should be computed explicitly. Efficient implementations determine the matrices $\mathrm{A}^\ell, \mathrm{B}^\ell, \mathrm{C}^\ell$ directly from

(approximate) low-rank factors of P, Q. With the advance of numerical algorithms for solving Lyapunov equations, see [16, 81] for recent surveys, nowadays these techniques can be applied to any kind of system for which linear systems $\mathrm{A}x = b$ with the matrix A as above are solvable, see [11] for details,

There also exist numerous variants of BT that can be useful as model reduction techniques in PDE-constrained optimization by exchanging P, Q by other useful pairs of positive (semi-)definite matrices, see [3, 11] and [14, Chapter 1] for some of these.

4.3. Applications to PDE-Constrained Optimization

The system-theoretic model reduction methods discussed in the previous sections have so far found wide-reaching applications in simulation and control, see, e.g., [3, 10, 11, 14, 28] and many more references therein. There have also been numerous attempts to apply balanced truncation directly to PDE control problems, mostly with the target of feedback control rather than optimization. The theory for balanced truncation of distributed parameter systems, i.e., linear instationary PDE control problems, was already derived in the 1980ies, see [32]. The use of this approach for deriving robust control strategies from the reduced-order model is discussed in [23]. The practical use of these approaches needs, of course, computational methods, and thus discretization at some stage of the reduction process. In [12], it is discussed how to combine BT with classical finite element discretization in the spatial variables of linear parabolic control problems while [70] goes a step further and derives an implementation of a balanced truncation algorithm where the discretization is delayed until the inner loop of the numerical algorithm is entered — the latter approach is more in the spirit of "optimize-than-discretize", while the approach in [12] can be seen as "semi-discretize-than-optimize". Balanced truncation and related methods have also been applied to flow control problems, see, e.g., [6, 13, 43, 91].

The methods based on rational interpolation have merely been used so far in simulation of dynamical systems, but also for feedback control purposes. Further reading on this includes [3, 7, 10, 14, 28, 29] and references therein. Their use for model reduction of infinite-dimensional control problems is discussed in [69] and for linear-quadratic parabolic optimal control problems in [90]. The recent paper [18] uses rational tangential interpolation for a flow control problem. Certainly, these methods can be employed in a similar fashion as balanced truncation in PDE-constrained optimization problems, but so far this topic has received less attention.

Note that all these approaches do not discretize in time, no discretization of the controls is needed whenever the control variables are only time-dependent, and only the spatial part needs to be discretized for controls of the form $u(\boldsymbol{x}, t) = v(\boldsymbol{x})\mathrm{u}(t)$. Hence, optimization based on reduced-order models obtained from these methods can be performed with respect to the original U_{ad} rather than a discretized version or subset of the full space of admissible controls. The full potential of this advantageous property has yet to be explored in PDE-constrained optimization. Little work has been dedicated

so far to using system-theoretic methods in an optimization context. The application of balanced truncation to a fully discrete linear-quadratic PDE optimal control problem is discussed in [9, Section 6.2]. The use of balanced truncation and rational interpolation techniques as well as POD for linear-quadratic parabolic optimal control problems is considered in [90]. Balanced truncation to accelerate a descent method for optimization of nonlinear evolution problems is analyzed in [25]. There, balanced truncation is applied to the linear adjoint system while the full state equation is solved. Speed-ups of factors 2–3 can be obtained this way compared to applying the same descent method to the full order optimality system. More research in the direction of using system-theoretic methods in a PDE-constrained optimization setting for evolution problems is certainly needed to leverage their full potential.

5. Conclusions

In this chapter, we have reviewed model reduction techniques for PDE-constrained optimization, where we have focused on instationary PDEs only. There is also a vast amount of literature on optimization of stationary PDEs, like linear elliptic PDEs in the simplest case. Model reduction methods as discussed here are not suitable for these problems as they rely on techniques for time-varying systems. Moreover, the computational complexity of instationary PDE-constrained optimization problems is often not so high that such a great benefit can be expected from model order reduction as for instationary problems. Nevertheless, model reduction techniques for such problems exist. These are based, e.g., on POD w.r.t. the optimization parameters rather than time, or on the Reduced Basis Method, see, e.g., [34, 52, 63, 68, 85] for some recent work on this.

We would also like to point out that there are many further possibilities for research. As already mentioned, the capabilities of system-theoretic methods in the context of PDE-constrained optimization are yet to be explored in detail. Also, the combined optimization with regard to a time-dependent control function plus one or more stationary design parameters as it occurs in many practical engineering applications is a widely open field.

Acknowledgment

Our thanks go to Günter Leugering for his well-structured and focused leadership throughout the six years of existence of the DFG Priority Program 1253 “Optimization with Partial Differential Equations”.

References

[1] K. Afanasiev and M. Hinze, *Adaptive control of a wake flow using proper orthogonal decomposition*. Lecture Notes in Pure and Applied Mathematics, 216:317–332, 2001.

[2] A. Alla and S. Volkwein, *Asymptotic stability of POD based model predictive control for a semilinear parabolic PDE*. Submitted, 2013. `http://kops.ub.uni-konstanz.de/handle/urn:nbn:de:bsz:352-253400`

[3] A.C. Antoulas, *Approximation of Large-Scale Dynamical Systems*, SIAM, Philadelphia, 2005.

[4] E. Arian, M. Fahl, and E.W. Sachs, *Trust-region proper orthogonal decomposition for flow control*. Technical Report 2000-25, ICASE, 2000.

[5] P. Astrid, S. Weiland, K. Willcox, and T. Backx, *Missing point estimation in models described by proper orthogonal decomposition*. IEEE Transactions on Automatic Control, 53(10):2237–2251, 2008.

[6] J.A. Atwell, J.T. Borggaard, and B.B. King, *Reduced-order controllers for Burgers' equation with a nonlinear observer*. International Journal of Applied Mathematics and Computer Science, 11(6):1311–1330, 2001.

[7] Z. Bai, *Krylov subspace techniques for reduced-order modeling of large-scale dynamical systems*. Applied Numerical Mathematics, 43:9–44, 2002.

[8] H.T. Banks, M.L. Joyner, B. Winchesky, and W.P. Winfree, *Nondestructive evaluation using a reduced-order computational methodology*. Inverse Problems, 16:1–17, 2000.

[9] U. Baur, *Control-Oriented Model Reduction for Parabolic Systems*, Dissertation, TU Berlin, 2008.

[10] U. Baur, P. Benner, and L. Feng, *Model order reduction for linear and nonlinear systems: a system-theoretic perspective*. Archives of Computational Methods in Engineering, to appear. Preliminary version available as Max Planck Institute Magdeburg Preprints MPIMD/14-07, March 2014.

[11] P. Benner, *System-theoretic methods for model reduction of large-scale systems: Simulation, control, and inverse problems*. In I. Troch and F. Breitenecker, eds., Proceedings of MathMod 2009, Vienna, February 11–13, vol. 35 of ARGESIM Report, 2009, pp. 126–145, 2009.

[12] P. Benner, *Balancing-related model reduction for parabolic control systems*. Proceedings of the 1st IFAC Workshop on Control of Systems Governed by Partial Differential Equations, pp. 257–262, IFAC Papers Online, 2013

[13] P. Benner and J. Heiland, *LQG-balanced truncation low-order controller for stabilization of laminar flows*. Max Planck Institute Magdeburg Preprints MPIMD/14-07, March 2014.

[14] P. Benner, V. Mehrmann, and D.C. Sorensen, *Dimension Reduction of Large-Scale Systems*, Lecture Notes in Computational Science and Engineering, vol. 45, Springer-Verlag, 2005.

[15] P. Benner, E.S. Quintana-Ortí, and G. Quintana-Ortí. *Balanced truncation model reduction of large-scale dense systems on parallel computers*. Mathematical and Computer Modelling of Dynamical Systems, 6(4):383–405, 2000.

[16] P. Benner and J. Saak. *Numerical solution of large and sparse continuous time algebraic matrix Riccati and Lyapunov equations: A state of the art survey*. GAMM Mitteilungen, 36(1):32–52, 2013.

[17] M. Bergmann and L. Cordier, *Optimal control of the cylinder wake in the laminar regime by trust-region methods and {POD} reduced-order models*, Journal of Computational Physics, 227(16):7813–7840, 2008.

[18] J.T. Borggaard and S. Gugercin, *Model reduction for DAEs with an application to flow control*. Department of Mathematics, Virginia Tech, Blacksburg, VA, 2014.

[19] E. Casas, *Pontryagin's principle for state-constrained boundary control problems of semilinear parabolic equations*. SIAM J. Control and Optimization, 35:1297–1327, 1997.

[20] D. Chapelle, A. Gariah, and J. Saint-Marie, *Galerkin approximation with proper orthogonal decomposition: new error estimates and illustrative examples*. ESAIM: Mathematical Modelling and Numerical Analysis, 46:731–757, 2012.

[21] X. Chen, S. Akella, and I.M. Navon, *A dual-weighted trust-region adaptive POD 4-D VAR applied to a finite-volume shallow water equations model on the sphere*, International Journal for Numerical Methods in Fluids, 68(3):377–402, 2012.

[22] R. Cont, N. Lantos, and O. Pironneau, *A reduced basis for option pricing*, SIAM Journal on Financial Mathematics, 2(1):287–316, 2011.

[23] R.F. Curtain, *Model reduction for control design for distributed parameter systems*. In Research directions in distributed parameter systems, volume 27 of Frontiers Appl. Math., 95–121. SIAM, Philadelphia, PA, 2003.

[24] L. Dedé, *Reduced basis method and a posteriori error estimation for parametrized linear-quadratic optimal control problems*. SIAM Journal on Scientific Computing, 32:997–1019, 2010.

[25] J.C. De Los Reyes and T. Stykel, *A balanced truncation based strategy for the optimal control of evolution problems*. Optimization Methods and Software, 26(4-5):673–694, 2011.

[26] M. Dihlmann and B. Haasdonk, *Certified nonlinear parameter optimization with reduced basis surrogate models*. Submitted, 2013.

[27] A.L. Dontchev, W.W. Hager, A.B. Poore, and B. Yang, *Optimality, stability, and convergence in nonlinear control*. Applied Mathematics and Optimization, 31:297–326, 1995.

[28] P. Feldmann and R.W. Freund, *Efficient linear circuit analysis by Padé approximation via the Lanczos process*. IEEE Transaction on Computer-Aided Design, 14:639–649, 1995.

[29] R. Freund, *Model reduction methods based on Krylov subspaces*. Acta Numerica 12:267–319, 2003.

[30] K. Gallivan, E. Grimme, and P. Van Dooren, *Asymptotic waveform evaluation via the Lanczos method*. Applied Mathematics Letters, 7:75–80, 1994.

[31] J. Ghiglieri and S. Ulbrich, *Optimal flow control based on POD and MPC and an application to the cancellation of Tollmien-Schlichting waves*. Optimization Methods and Software, to appear.

[32] K. Glover, R.F. Curtain, and J.R. Partington, *Realisation and approximation of linear infinite-dimensional systems with error bounds*. SIAM Journal on Control and Optimization, 26(4):863–898, 1988.

[33] C.-C. Gräßle, S. Rogg, S. Trenz, and S. Volkwein, *Adaptive trust region POD with a-posteriori error estimation for nonlinear optimal control*. In preparation, 2014.

[34] M.A. Grepl and M. Kärcher, *Reduced basis a posteriori error bounds for parametrized linear-quadratic elliptic optimal control problems.* Comptes Rendus de l'Acadmie des Sciences — Series I, 349:873–877, 2011.

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

[36] E. Grimm, *Optimality system POD and a-posteriori error analysis for linear-quadratic optimal control problems.* Master thesis, University of Konstanz, 2013.

[37] M. Gubisch and S. Volkwein, *Proper orthogonal decomposition for linear-quadratic optimal control.* Submitted, 2013. `http://kops.ub.uni-konstanz.de/handle/urn:nbn:de:bsz:352-250378`

[38] M. Gubisch and S. Volkwein, *POD a-posteriori error analysis for optimal control problems with mixed control-state constraints.* Computational Optimization and Applications, to appear.

[39] S. Gugercin, *Projection methods for model reduction of large-scale dynamical systems.* Ph.D. thesis, Rice University, Houston, 2003.

[40] S. Gugercin, A.C. Antoulas, and C. Beattie, $\mathcal{H}_2$ *Model reduction for large-scale dynamical systems.* SIAM Journal on Matrix Analysis and Applications, 30(2):609–638, 2008.

[41] B. Haasdonk, *Convergence rates of the POD-Greedy method.* ESAIM: Mathematical Modelling and Numerical Analysis, 47:859–873, 2013.

[42] B. Haasdonk and M. Ohlberger. *Reduced basis method for finite volume approximations of parametrized linear evolution equations.* ESAIM: Mathematical Modelling and Numerical Analysis, 42:277–302, 2008.

[43] M. Heinkenschloss, D.C. Sorensen, and K. Sun. *Balanced truncation model reduction for a class of descriptor systems with application to the Oseen equations.* SIAM Journal on Scientific Computing, 30:1038–1063, 2008.

[44] C. Himpe and M. Ohlberger. *Cross-gramian based combined state and parameter reduction.* Submitted, 2013.

[45] M. Hintermüller, K. Ito, and K. Kunisch, *The primal-dual active set strategy as a semismooth Newton method.* SIAM Journal on Optimization, 13:865–888, 2003.

[46] M. Hinze, R. Pinnau, M. Ulbrich, and M. Ulbrich, *Optimization with PDE Constraints*, Mathematical Modeling: Theory and Applications, vol. 23, Springer, Berlin, 2009.

[47] M. Hinze and F. Tröltzsch, *Discrete concepts versus error analysis in PDE constrained optimization.* GAMM Mitteilungen, 33:148–162, 2010.

[48] M. Hinze and S. Volkwein, *Proper orthogonal decomposition surrogate models for nonlinear dynamical systems: error estimates and suboptimal control.* Chapter 10 (pages 261–306) of [14].

[49] M. Hinze and S. Volkwein, *Error estimates for abstract linear-quadratic optimal control problems using proper orthogonal decomposition.* Computational Optimization and Applications, 39:319–345, 2008.

[50] P. Holmes, J.L. Lumley, G. Berkooz, and C.W. Rowley, *Turbulence, Coherent Structures, Dynamical Systems and Symmetry.* Cambridge Monographs on Mechanics, Cambridge University Press, second edition, 2012.

[51] K. Ito and S.S. Ravindran, *A reduced basis method for control problems governed by PDEs*. In W. Desch, F. Kappel, and K. Kunisch, eds., Control and Estimation of Distributed Parameter Systems. Proceedings of the International Conference in Vorau, 1996, Birkhäuser-Verlag, Basel, 126:153–168, 1998.

[52] M. Kahlbacher and S. Volkwein, *POD a-posteriori error based inexact SQP method for bilinear elliptic optimal control problems*. ESAIM: Mathememtical Modelling and Numerical Analysis, 46:491–511, 2012.

[53] E. Kammann, F. Tröltzsch, and S. Volkwein,*A method of a-posteriori error estimation with application to proper orthogonal decomposition*, ESAIM: Mathematical Modelling and Numerical Analysis, 47:555–581, 2013.

[54] K. Kunisch and S. Volkwein, *Proper orthogonal decomposition for optimality systems*. ESAIM: Mathematical Modelling and Numerical Analysis, 42:1–23, 2008.

[55] K. Kunisch, S. Volkwein, and L. Xie, *HJB-POD based feedback design for the optimal control of evolution problems*. SIAM Journal on Applied Dynamical Systems, 3:701–722, 2004.

[56] S. Lall, J.E. Marsden, and S. Glavaski. *A subspace approach to balanced truncation for model reduction of nonlinear control systems*. International Journal of Robust and Nonlinear Control, 12:519–535, 2002.

[57] O. Lass and S. Volkwein, *Parameter identification for nonlinear elliptic-parabolic systems with application in lithium-ion battery modeling*. Submitted, 2013. `http://kops.ub.uni-konstanz.de/handle/urn:nbn:de:bsz:352-253400`

[58] F. Leibfritz and S. Volkwein, *Reduced order output feedback control design for PDE systems using proper orthogonal decomposition and nonlinear semidefinite programming*. Linear Algebra and Its Applications, 415:542–575, 2006.

[59] S. Li, L. Feng, P. Benner, and A. Seidel-Morgensterna, *Using surrogate models for efficient optimization of simulated moving bed chromatography*. Computers and Chemical Engineering, 2014. DOI: 10.1016/j.compchemeng.2014.03.024.

[60] H.V. Ly and H.T. Tran, *Modeling and control of physical processes using proper orthogonal decomposition*. Mathematical and Computer Modeling, 33:223–236, 2001.

[61] V. Mehrmann and T. Stykel, *Balanced truncation model reduction for large-scale systems in descriptor form*. Chapter 3 (pages 83–115) of [14].

[62] M. Müller, *Uniform Convergence of the POD Method and Applications to Optimal Control*. Ph.D thesis, University of Graz, 2011.

[63] F. Negri, G. Rozza, A. Manzoni, and A. Quarteroni, *Reduced basis method for parametrized elliptic optimal control problems*. SIAM Journal on Scientific Computing, 35(5):A2316–A2340, 2013.

[64] B. Noack, K. Afanasiev, M. Morzynski, G. Tadmor, and F. Thiele, *A hierarchy of low-dimensional models for the transient and post-transient cylinder wake*. Journal of Fluid. Mechanics, 497:335–363, 2003.

[65] J. Nocedal and S. J. Wright, *Numerical Optimization*, Springer Series in Operations Research and Financial Engineering, second edition, 2006.

[66] 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, 2006.

[67] O. Pironneau. *Calibration of options on a reduced basis.* Journal of Computational and Applied Mathematics , 232:139–147, 2009.

[68] M. Ohlberger and M. Schäfer, *Error control based model reduction for parameter optimization of elliptic homogenization problems.* Proceedings of the 1st IFAC Workshop on Control of Systems Governed by Partial Differential Equations, pp. 251–256, IFAC Papers Online, 2013.

[69] M.R. Opmeer, *Model order reduction by balanced proper orthogonal decomposition and by rational interpolation.* IEEE Transactions on Automatic Control, 57(2):472–477, 2012.

[70] M.R. Opmeer, T. Reis, and W. Wollner, *Finite-rank ADI iteration for operator Lyapunov equations.* SIAM Journal on Control and Optimization, 51(5):4084–4117, 2013.

[71] M. Rathinam and L. Petzold, *Dynamic iteration using reduced order models: a method for simulation of large scale modular systems.* SIAM Journal on Numerical Analysis, 40:1446–1474, 2002.

[72] S.S. Ravindran, *Reduced-order adaptive controllers for fluid flows using POD.* SIAM Journal on Scintific Computing, 15:457?478, 2000.

[73] J.-P. Raymond and H. Zidani, *Hamiltonian Pontryagin's principle for state-constrained control problems governed by semilinear parabolic equations.* Applied Mathematics and Optimization, 39:143–177, 1999.

[74] M. Reed and B. Simon, *Methods of Modern Mathematical Physics I: Functional Analysis.* Academic Press, New York, 1980.

[75] C.W. Rowley, *Model reduction for fluids, using balanced proper orthogonal decomposition.* International Journal on Bifurcation and Chaos, 15:997–1013, 2005.

[76] E.W. Sachs and S. Volkwein, *POD Galerkin approximations in PDE-constrained optimization*, GAMM Mitteilungen, 33:194–208, 2010.

[77] A. Schiela and M. Weiser, *Superlinear convergence of the control reduced interior point method for PDE constrained optimization.* Computational Optimization and Applications, 39:369–393, 2008.

[78] W.H.A. Schilders, H.A. van der Vorst, and J. Rommes, *Model Order Reduction: Theory, Research Aspects and Applications*, Mathematics in Industry, vol. 13, Springer, Berlin, 2008.

[79] E.W. Sachs and M. Schneider, *Reduced order models for the implied variance und local volatility*, Technical report, Trier University, 2014.

[80] E.W. Sachs and M. Schu, *A-priori error estimates for reduced order models in finance.* ESAIM: Mathematical Modelling and Numerical Analysis, 47:449–469, 2013.

[81] V. Simoncini. *Computational methods for linear matrix equations.* Preprint, March 2013. `http://www.dm.unibo.it/œsimoncin/`.

[82] J.R. Singler, *New POD expressions, error bounds, and asymptotic results for reduced order models of parabolic PDEs.* SIAM Journal on Numerical Analysis, 52:852–876, 2014.

[83] L. Sirovich, *Turbulence and the dynamics of coherent structures. Parts I-II.* Quarterly of Applied Mathematics, XVL:561–590, 1987.

[84] A. Studinger and S. Volkwein, *Numerical analysis of POD a-posteriori error estimation for optimal control*. International Series of Numerical Mathematics, 164:137–158, 2013

[85] T. Tonn, K. Urban, and S. Volkwein, *Comparison of the reduced-basis and POD a-posteriori error estimators for an elliptic linear quadratic optimal control problem*. Mathematical and Computer Modelling of Dynamical Systems, 17:355–369, 2011.

[86] F. Tröltzsch, *Optimal Control of Partial Differential Equations*, American Mathematical Society, Graduate Studies Mathematics, vol. 112, Springer, New York, 2010.

[87] F. Tröltzsch and S. Volkwein, *POD a-posteriori error estimates for linear-quadratic optimal control problems*. Computational Optimization and Applications, 44:83–115, 2009.

[88] M. Ulbrich and S.Ulbrich, *Primal-dual interior-point methods for PDE-constrained optimization*. Mathematical Programming, 117: 435–485, 2008 .

[89] S. Volkwein, *Optimality system POD and a-posteriori error analysis for linear-quadratic problems*. Control and Cybernetics, 40:1109–1125, 2011.

[90] G. Vossen and S. Volkwein, *Model reduction techniques with a-posteriori error analysis for linear-quadratic optimal control problems*. Numerical Algebra, Control and Optimization, 2:465–485, 2012.

[91] K. Willcox and J. Peraire, *Balanced model reduction via the proper orthogonal decomposition*. American Institute of Aeronautics and Astronautics (AIAA), 2323–2330, 2002.

[92] Y. Yue and K. Meerbergen. *Accelerating optimization of parametric linear systems by model order reduction*. SIAM Journal on Optimization, 23(2):1344–1370, 2013.

[93] K. Zhou, J.C. Doyle, and K. Glover, *Robust and Optimal Control*. Prentice-Hall, Upper Saddle River, NJ, 1996.

Peter Benner
Max Planck Institute for Dynamics of Complex Technical Systems
Sandtorstraße 1
39106 Magdeburg
Germany
e-mail: benner@mpi-magdeburg.mpg.de

Ekkehard Sachs
FB 4 – Mathematik
University of Trier
54286 Trier
Germany
e-mail: sachs@uni-trier.de

Stefan Volkwein
Department of Mathematics and Statistics
University of Konstanz
Universitätsstraße 10
78464 Konstanz
Germany
e-mail: stefan.volkwein@uni-konstanz.de