SPACE-TIME GALERKIN POD WITH APPLICATION IN OPTIMAL CONTROL OF SEMI-LINEAR PARABOLIC PARTIAL DIFFERENTIAL EQUATIONS

MANUEL BAUMANN*, PETER BENNER†, AND JAN HEILAND†

Abstract. In the context of Galerkin discretizations of a partial differential equation (PDE), the modes of the classical method of Proper Orthogonal Decomposition (POD) can be interpreted as the ansatz and trial functions of a low-dimensional Galerkin scheme. If one also considers a Galerkin method for the time integration, one can similarly define a POD reduction of the temporal component. This has been described earlier but not expanded upon – probably because the reduced time discretization globalizes time which is computationally inefficient. However, in finite-time optimal control systems, time *is* a global variable and there is no disadvantage from using a POD reduced Galerkin scheme in time. In this paper, we provide a newly developed generalized theory for space-time Galerkin POD, prove its optimality in the relevant function spaces, show its application for the optimal control of nonlinear PDEs, and, by means of a numerical example with Burgers' equation, discuss the competitiveness by comparing to standard approaches.

1. Introduction. The method of *Proper Orthogonal Decomposition* (POD) is a standard model reduction tool. For a generic dynamical system

$$\dot{v} = f(t, v), \tag{1}$$

on the time interval $(0, T]$ with a solution v with $v(t) \in \mathbb{R}^N$ and using samples $v(t_j)$, POD provides a set of $\hat{n}$ so-called POD modes $\hat{v}_1$, ..., $\hat{v}_{\hat{n}} \in \mathbb{R}^N$ which optimally parametrize the solution trajectory. As a result, the system (1) can be projected down to a system of reduced spatial dimension $\hat{n}$ that often reflects the dynamical behavior of (1) well. If the considered system stems from a *Finite Element* (FEM) discretization of a PDE, then the modes $\hat{v}_i$, $i = 1, \dots, \hat{n}$, can be interpreted as ansatz functions in the finite element space $\mathcal{Y}$ and the projected system as a particular Galerkin projection of the underlying PDE.

In this paper we provide a theoretical framework and show cases for a space-time Galerkin POD method. Some of the underlying ideas for this generalization of POD have been developed and tested in our earlier works [2, 3].

The first innovation of the proposed generalized POD approach bases on the observation that instead of the discrete time samples $v(t_j)$, one may use the projection of v onto the finite dimensional subspace $\mathcal{S} \cdot \mathcal{Y}$, where $\mathcal{S}$ is a, say, k-dimensional subspace of $L^2(0, T)$. The second innovation is that the projection onto $\mathcal{S} \cdot \mathcal{Y}$ can be interpreted as Galerkin discretization in time which can be reduced analoguously to the POD reduction of the space dimension. The resulting scheme is a POD reduced space-time Galerkin discretization.

This basic idea of a space-time POD has already been taken up in [19], but not progressed since then. We think that this is due to the fact that temporal POD destructs the causality in time which makes it very inefficient for numerical simulations. In fact, the POD reduced time ansatz functions are global such that the space-time Galerkin system has to be solved as a whole rather than in sequences of time slobs as in standard time-stepping or discontinuous Galerkin schemes [14, 17]. Thus, the reduced space-time scheme cannot compete with, e.g., a spatial POD combined with a standard Runge-Kutta solver. However, in finite-time optimal control problems, the

*Delft Institute of Applied Mathematics (m.m.baumann@tudelft.nl),
†Max Planck Institute for Dynamics of Complex Technical Systems Magdeburg (benner,heiland@mpi-magdeburg.mpg.de),

time *is* a global variable and, as we will show by numerical examples, the space-time Galerkin discretization becomes very competitive.

The need and the potential of also reducing the time dimension of a reduced order model have been discussed in [5]. There – similar to our observation that an SVD of a matrix of measurements also reveals compressed time information – it is proposed to use the right singular vectors of a classical snapshot matrix for forecasting.

We want to point out that the method of *Proper Generalized Decomposition* (PGD) is related to the proposed space-time Galerkin POD only in so far as for PGD also space-time (and parameter) tensor bases are used for the modelling; see, e.g., [6]. However, the PGD approach seeks to successively build up the bases by collocation, *greedy algorithms*, and fixed-point iteration, whereas our approach reduces a given basis on the base of measurements. For the same reasons, the connection of the presented approach to other tensor-based low-dimensional approximation schemes [10, 15] as well as to *Reduced Basis* approaches [20] is only marginal.

This paper is organized as follows: At first, we introduce the mathematical framework and rigorously prove the optimality of the reduced space and time bases. Then we illustrate how the reduced bases can be used for low-dimensional space-time Galerkin approximations. In particular, we address how to treat quadratic nonlinearities, how to incorporate initial and terminal values, and how to set up the bases for a general PDE by means of standard approximation schemes. Finally, we illustrate the performance of the space-time Galerkin POD approach for the optimal control of Burgers' equation and compare it to well-established gradient-based methods combined with standard POD.

2. Space-Time Galerkin POD. In this section, we provide the analytical framework for space-time POD. We introduce the considered function spaces and directly prove the optimality of the POD projection in the respective space-time L^2 norm. For a time interval $(0, T)$ and a spatial domain Ω, consider the space-time function space $L^2(0, T; L^2(\Omega))$. Let

$$\mathcal{S} = \operatorname{span}\{\psi_1, \dots, \psi_s\} \subset L^2(0, T) \quad \text{and} \quad \mathcal{Y} = \operatorname{span}\{\nu_1, \dots, \nu_q\} \subset L^2(\Omega)$$

be finite dimensional subspaces of dimension s and q, respectively, and let

$$\mathcal{X} = \mathcal{S} \cdot \mathcal{Y} \subset L^2(0, T; L^2(\Omega)). \tag{2}$$

The space-time L^2-orthogonal projection $\bar{x} := \Pi_{\mathcal{S}\cdot\mathcal{Y}} x$ of some $x \in L^2(0, T; L^2(\Omega))$ onto $\mathcal{X}$ is given as

$$\bar{x}(\xi, \tau) = \sum_{j=1}^{s} \sum_{i=1}^{q} \mathbf{x}_{i,j} \nu_i(\xi) \psi_j(\tau), \tag{3}$$

where the coefficients $\mathbf{x}_{i,j}$ are the entries of the matrix

$$\mathbf{X} = \left[\mathbf{x}_{i,j}\right]_{i=1,\dots,q}^{j=1,\dots,s} := \mathbf{M}_{\mathcal{Y}}^{-1} \begin{bmatrix} ((x, \nu_1\psi_1))_{\mathcal{S}\cdot\mathcal{Y}} & \dots & ((x, \nu_1\psi_s))_{\mathcal{S}\cdot\mathcal{Y}} \\ \vdots & \ddots & \vdots \\ ((x, \nu_q\psi_1))_{\mathcal{S}\cdot\mathcal{Y}} & \dots & ((x, \nu_q\psi_s))_{\mathcal{S}\cdot\mathcal{Y}} \end{bmatrix} \mathbf{M}_{\mathcal{S}}^{-1}, \tag{4}$$

where

$$((x, \nu_i\psi_j))_{\mathcal{S}\cdot\mathcal{Y}} := ((x, \nu_i)_{\mathcal{Y}}, \psi_j)_{\mathcal{S}} := \int_0^T \Big(\int_\Omega x(\xi, \tau) \nu_i(\xi) \,\mathrm{d}\xi \Big) \psi_j(\tau) \,\mathrm{d}\tau.$$

Here, $\mathbf{M}_{\mathcal{Y}}^{-1}$ and $\mathbf{M}_{\mathcal{S}}^{-1}$ are the inverses of the mass matrices with respect to space and time,

$$\mathbf{M}_{\mathcal{Y}} := \left[(\nu_i, \nu_j)_{\mathcal{Y}}\right]_{i=1,\dots,q}^{j=1,\dots,q} \quad \text{and} \quad \mathbf{M}_{\mathcal{S}} := \left[(\psi_i, \psi_j)_{\mathcal{S}}\right]_{i=1,\dots,s}^{j=1,\dots,s}. \tag{5}$$

REMARK 2.1. *We will refer to $\mathcal{X} = \mathcal{S} \cdot \mathcal{Y}$ as the measurement space, to the basis functions of $\mathcal{Y}$ and $\mathcal{S}$ as measurement functions, and to $\mathbf{X}$ as the measurement matrix. This means that a function in $L^2(0, T; L^2(\Omega))$ can be measured in $\mathcal{X}$, e.g. via its projection onto $\mathcal{X}$, and, the other way around, an element $\mathbf{X}$ of $\mathcal{X}$ can be seen as a measurement of some functions in $L^2(0, T; L^2(\Omega))$.*

We introduce some representations of the inner product and the norm of functions in $\mathcal{S} \cdot \mathcal{Y}$.

LEMMA 2.2 (Space-time discrete L^2-product). *Let*

$$x^1 = \sum_{j=1}^{s} \sum_{i=1}^{q} \mathbf{x}_{i,j}^1 \nu_i \psi_j \in \mathcal{S} \cdot \mathcal{Y}, \quad x^2 = \sum_{j=1}^{s} \sum_{i=1}^{q} \mathbf{x}_{i,j}^2 \nu_i \psi_j \in \mathcal{S} \cdot \mathcal{Y},$$

then, with

$$\mathbf{x}^\ell = [\mathbf{x}_{1,1}^\ell, \dots, \mathbf{x}_{q,1}^\ell, \mathbf{x}_{1,2}^\ell, \dots, \mathbf{x}_{q,2}^\ell, \dots, \mathbf{x}_{1,s}^\ell, \dots, \mathbf{x}_{q,s}^\ell]^\mathsf{T} =: \operatorname{vec}(\mathbf{X}^\ell), \quad \ell = 1, 2,$$

the inner product in $\mathcal{S} \cdot \mathcal{Y}$ is given as

$$((x^1, x^2))_{\mathcal{S} \cdot \mathcal{Y}} = \int_0^T \int_\Omega x^1 x^2 \, \mathrm{d}\xi \, \mathrm{d}\tau = (\mathbf{x}^1)^\mathsf{T} (\mathbf{M}_{\mathcal{S}} \otimes \mathbf{M}_{\mathcal{Y}}) \, \mathbf{x}^2 \tag{6}$$

and the induced norm as

$$\|x^\ell\|_{\mathcal{S} \cdot \mathcal{Y}}^2 := ((x^\ell, x^\ell))_{\mathcal{S} \cdot \mathcal{Y}} = \|\mathbf{x}^\ell\|_{\mathbf{M}_{\mathcal{S}} \otimes \mathbf{M}_{\mathcal{Y}}}^2 = \|\mathbf{M}_{\mathcal{Y}}^{1/2} \mathbf{X}^\ell \mathbf{M}_{\mathcal{S}}^{1/2}\|_F^2, \quad \ell = 1, 2, \tag{7}$$

where $\|\cdot\|_{\mathbf{M}_{\mathcal{S}} \otimes \mathbf{M}_{\mathcal{Y}}}$ denotes the Euclidean vector norm weighted by $\mathbf{M}_{\mathcal{S}} \otimes \mathbf{M}_{\mathcal{Y}}$, and $\|\cdot\|_F$ is the Frobenius norm.

Proof. Straight-forward calculations. ∎

REMARK 2.3. *In practical applications, one uses a* Cholesky factorization *of the mass matrices* (5) *rather than the square-root.*

COROLLARY 2.4. *Let $\mathbf{M}_{\mathcal{S}} = \mathbf{L}_{\mathcal{S}} \mathbf{L}_{\mathcal{S}}^\mathsf{T}$ and $\mathbf{M}_{\mathcal{Y}} = \mathbf{L}_{\mathcal{Y}} \mathbf{L}_{\mathcal{Y}}^\mathsf{T}$ be given in factored form. Then, for a given $x \in \mathcal{S} \cdot \mathcal{Y}$ with its coefficient matrix $\mathbf{X}$ and vector $\mathbf{x} = \operatorname{vec}(\mathbf{X})$ it holds that*

$$\|x\|_{\mathcal{S} \cdot \mathcal{Y}}^2 = \|\mathbf{x}\|_{\mathbf{M}_{\mathcal{S}} \otimes \mathbf{M}_{\mathcal{Y}}}^2 = \|\mathbf{L}_{\mathcal{Y}}^\mathsf{T} \mathbf{X} \mathbf{L}_{\mathcal{S}}\|_F^2.$$

Proof.

$$\begin{aligned} \|\mathbf{x}\|_{\mathbf{M}_{\mathcal{S}} \otimes \mathbf{M}_{\mathcal{Y}}}^2 &= \mathbf{x}^\mathsf{T} (\mathbf{M}_{\mathcal{S}} \otimes \mathbf{M}_{\mathcal{Y}}) \mathbf{x} = \mathbf{x}^\mathsf{T} (\mathbf{L}_{\mathcal{S}} \otimes \mathbf{L}_{\mathcal{Y}}) \cdot (\mathbf{L}_{\mathcal{S}}^\mathsf{T} \otimes \mathbf{L}_{\mathcal{Y}}^\mathsf{T}) \mathbf{x} \\ &= \|(\mathbf{L}_{\mathcal{S}}^\mathsf{T} \otimes \mathbf{L}_{\mathcal{Y}}^\mathsf{T}) \mathbf{x}\|_2^2 = \|\operatorname{vec}(\mathbf{L}_{\mathcal{Y}}^\mathsf{T} \mathbf{X} \mathbf{L}_{\mathcal{S}})\|_2^2 = \|\mathbf{L}_{\mathcal{Y}}^\mathsf{T} \mathbf{X} \mathbf{L}_{\mathcal{S}}\|_F^2, \end{aligned}$$

as it follows from basic properties and relations between the Kronecker product, the vectorization operator, and the Frobenius norm. ∎

From now on, we will always consider the factorized form. In theory, one can always replace the factors by the square roots of the respective mass matrices.

Next, we will consider a given function $x \in \mathcal{S} \cdot \mathcal{Y}$ and determine low-dimensional subspaces of $\mathcal{Y}$ and $\mathcal{S}$ that can provide low-dimensional approximations to x in a norm-optimal way.

LEMMA 2.5 (Optimal low-rank bases in space). *Given $x \in \mathcal{S} \cdot \mathcal{Y}$ and the associated matrix of coefficients $\mathbf{X}$. The best-approximating $\hat{q}$-dimensional subspace $\hat{\mathcal{Y}}$ in the sense that the projection error $\|x - \Pi_{\mathcal{S}\cdot\hat{\mathcal{Y}}} x\|_{\mathcal{S}\cdot\mathcal{Y}}$ is minimal over all subspaces of $\mathcal{Y}$ of dimension $\hat{q}$ is given as* $\operatorname{span}\{\hat{\nu}_i\}_{i=1,\dots,\hat{q}}$, *where*

$$\begin{bmatrix} \hat{\nu}_1 \\ \hat{\nu}_2 \\ \vdots \\ \hat{\nu}_{\hat{q}} \end{bmatrix} = V_{\hat{q}}^{\mathsf{T}} \mathbf{L}_{\mathcal{Y}}^{-1} \begin{bmatrix} \nu_1 \\ \nu_2 \\ \vdots \\ \nu_q \end{bmatrix}, \tag{8}$$

where $V_{\hat{q}}$ is the matrix of the $\hat{q}$ leading left singular vectors of the matrix

$$\mathbf{L}_{\mathcal{Y}}^{\mathsf{T}} \mathbf{X} \mathbf{L}_{\mathcal{S}}.$$

Proof. For the time dimension at fixed index j, we consider

$$y := \sum_{i=1}^{q} \mathbf{x}_{i,j} \nu_i = \begin{bmatrix} \mathbf{x}_{1,j} & \dots & \mathbf{x}_{q,j} \end{bmatrix} \begin{bmatrix} \nu_1 \\ \vdots \\ \nu_q \end{bmatrix} \in \mathcal{Y}.$$

Next, we determine the orthogonal projection of y onto $\hat{\mathcal{Y}}$. Therefore, we write y as a function in $\hat{\mathcal{Y}}$ and a remainder $\hat{R}$ in the orthogonal complement:

$$y = \begin{bmatrix} \mathbf{x}_{1,j} & \dots & \mathbf{x}_{q,j} \end{bmatrix} \begin{bmatrix} \nu_1 \\ \vdots \\ \nu_q \end{bmatrix} = \begin{bmatrix} \beta_1 & \dots & \beta_{\hat{q}} \end{bmatrix} \begin{bmatrix} \hat{\nu}_1 \\ \vdots \\ \hat{\nu}_{\hat{q}} \end{bmatrix} + \hat{R}.$$

We determine the coefficients β_k, $k = 1, \dots, \hat{q}$, by testing against the basis functions of $\hat{\mathcal{Y}}$. By mutual orthogonality of $\hat{\nu}_i$, $i = 1, \dots, \hat{q}$, and their orthogonality against $\hat{R}$, it follows that

$$\begin{aligned} \beta_k = (\sum_{i=1}^{\hat{q}} \beta_i \hat{\nu}_i, \hat{\nu}_k)_{\mathcal{Y}} = (\hat{R} + \sum_{i=1}^{\hat{q}} \beta_i \hat{\nu}_i, \hat{\nu}_k)_{\mathcal{Y}} &= (\sum_{i=1}^{q} \mathbf{x}_{i,j} \nu_i, \hat{\nu}_k)_{\mathcal{Y}} \\ &\overset{(*)}{=} \begin{bmatrix} \mathbf{x}_{i,j} & \dots & \mathbf{x}_{q,j} \end{bmatrix} \mathbf{M}_{\mathcal{Y}} \mathbf{L}_{\mathcal{Y}}^{-\mathsf{T}} V_{\hat{q},k} \\ &= \begin{bmatrix} \mathbf{x}_{i,j} & \dots & \mathbf{x}_{q,j} \end{bmatrix} \mathbf{L}_{\mathcal{Y}} V_{\hat{q},k}, \end{aligned}$$

where in $\overset{(*)}{=}$ we have used that $\hat{\nu}_k = \begin{bmatrix} \nu_1 & \dots & \nu_q \end{bmatrix} \mathbf{L}_{\mathcal{Y}}^{-\mathsf{T}} V_{\hat{q},k}$ and where $V_{\hat{q},k}$ is the k-th column of $V_{\hat{q}}$ in (8). Thus, we find that the coefficients of the orthogonal projection of y onto $\hat{\mathcal{Y}}$ in the bases of $\hat{\mathcal{Y}}$ and $\mathcal{Y}$ are given through

$$\begin{aligned}\hat{y} = \sum_{i=1}^{\hat{q}} \beta_i \hat{\nu}_i = \begin{bmatrix}\beta_1 & \dots & \beta_{\hat{q}}\end{bmatrix} \begin{bmatrix}\hat{\nu}_1 \\ \vdots \\ \hat{\nu}_{\hat{q}}\end{bmatrix} &= \begin{bmatrix}\mathbf{x}_{1,j} & \dots & \mathbf{x}_{q,j}\end{bmatrix} \mathbf{L}_{\mathcal{Y}} V_{\hat{q}} \begin{bmatrix}\hat{\nu}_1 \\ \vdots \\ \hat{\nu}_q\end{bmatrix} \\ &= \begin{bmatrix}\mathbf{x}_{1,j} & \dots & \mathbf{x}_{q,j}\end{bmatrix} \mathbf{L}_{\mathcal{Y}} V_{\hat{q}} V_{\hat{q}}^{\mathsf{T}} \mathbf{L}_{\mathcal{Y}}^{-1} \begin{bmatrix}\nu_1 \\ \vdots \\ \nu_q\end{bmatrix} \\ &=: \begin{bmatrix}\hat{\mathbf{x}}_{1,j} & \dots & \hat{\mathbf{x}}_{q,j}\end{bmatrix} \begin{bmatrix}\nu_1 \\ \vdots \\ \nu_q\end{bmatrix}.\end{aligned}$$

Noting that $\begin{bmatrix}\mathbf{x}_{1,j} & \dots & \mathbf{x}_{q,j}\end{bmatrix}^{\mathsf{T}}$ makes up the j-th column of the matrix $\mathbf{X}$ associated with x, we conclude that the matrix $\hat{\mathbf{X}}$ of coefficients associated with $\Pi_{\mathcal{S}\cdot\hat{\mathcal{Y}}}x$ is given as

$$\hat{\mathbf{X}} = \mathbf{L}_{\mathcal{Y}}^{-\mathsf{T}} V_{\hat{q}} V_{\hat{q}}^{\mathsf{T}} \mathbf{L}_{\mathcal{Y}}^{\mathsf{T}} \mathbf{X}$$

and, by Corollary 2.4, we have that

$$\begin{aligned}\|x - \Pi_{\mathcal{S}\cdot\hat{\mathcal{Y}}}x\|_{\mathcal{S}\cdot\mathcal{Y}} &= \|\mathbf{L}_{\mathcal{Y}}^{\mathsf{T}}\mathbf{X}\mathbf{L}_{\mathcal{S}} - \mathbf{L}_{\mathcal{Y}}^{\mathsf{T}}\hat{\mathbf{X}}\mathbf{L}_{\mathcal{S}}\|_F = \|\mathbf{L}_{\mathcal{Y}}^{\mathsf{T}}[\mathbf{X} - \hat{\mathbf{X}}]\mathbf{L}_{\mathcal{S}}\|_F \\ &= \|\mathbf{L}_{\mathcal{Y}}^{\mathsf{T}}\mathbf{X}\mathbf{L}_{\mathcal{S}} - V_{\hat{q}}V_{\hat{q}}^{\mathsf{T}}\mathbf{L}_{\mathcal{Y}}^{\mathsf{T}}\mathbf{X}\mathbf{L}_{\mathcal{S}}\|_F\end{aligned}$$

which is minimized over all $V_{\hat{q}} \in \mathbb{R}^{q,\hat{q}}$ matrices by taking $V_{\hat{q}}$ as the matrix of the $\hat{q}$ leading left singular vectors of $\mathbf{L}_{\mathcal{Y}}^{\mathsf{T}}\mathbf{X}\mathbf{L}_{\mathcal{S}}$. ∎

The same arguments apply to the transpose of $\mathbf{X}$:

LEMMA 2.6 (Optimal low-rank bases in time). *Given $x \in \mathcal{S}\cdot\mathcal{Y}$ and the associated matrix of coefficients $\mathbf{X}$. The best-approximating $\hat{s}$-dimensional subspace $\hat{\mathcal{S}}$ in the sense that the projection error $\|x - \Pi_{\hat{\mathcal{S}}\cdot\mathcal{Y}}x\|_{\mathcal{S}\cdot\mathcal{Y}}$ is minimal over all subspaces of $\mathcal{S}$ of dimension $\hat{s}$ is given as* $\operatorname{span}\{\hat{\psi}_j\}_{j=1,\dots,\hat{s}}$, *where*

$$\begin{bmatrix}\hat{\psi}_1 \\ \hat{\psi}_2 \\ \vdots \\ \hat{\psi}_{\hat{s}}\end{bmatrix} = U_{\hat{s}}^{\mathsf{T}} \mathbf{L}_{\mathcal{S}}^{-1} \begin{bmatrix}\psi_1 \\ \psi_2 \\ \vdots \\ \psi_s\end{bmatrix}, \tag{9}$$

where $U_{\hat{s}}$ is the matrix of the $\hat{s}$ leading right singular vectors of

$$\mathbf{L}_{\mathcal{Y}}^{\mathsf{T}}\mathbf{X}\mathbf{L}_{\mathcal{S}}.$$

REMARK 2.7. *The approximation results Lemma 2.5 and Lemma 2.6 hold in the space-time L^2 norm, which is the appropriate norm for the considered functions and which is not part of the standard POD approach. However, the need for the right norms have been accounted for through the use of weighted inner products or weighted sums. If one lets $\mathcal{S}$ degenerate to a set of Dirac deltas, then Lemma 2.5 reduces to the optimality result [18, Thm. 1.8] for the standard POD approximation in the sense that the inner product is weighted with the FEM mass matrix. If one chooses $\mathcal{S}$ such that the induced time Galerkin scheme resembles a time discretization by the trapezoidal* *rule (in fact, for any Runge-Kutta scheme and choice of discretization points there exists a corresponding (discontinuous) Galerkin scheme), then Lemma 2.5 reduces to the optimality conditions for the* continuous POD *approach given in [18, Sec. 1.3].*

REMARK 2.8. *The idea of generalized measurements also works as a generalization of POD for a reduction of the state space. Consider the dynamical system (1), and define* $\mathbf{X}_{\mathcal{S}} := [(v_i, \psi_j)_{\mathcal{S}}]_{i=1,\dots,q}^{j=1,\dots,s}$*, where* v_i *is the* i*-th component of the vector-valued solution. Then the leading left singular vectors of the matrix* $\mathbf{X}_{\mathcal{S}}\mathbf{L}_{\mathcal{S}}^{-1}$ *are generalized POD modes and a projection of (1) onto the space spanned by those modes yields a POD-reduced dynamical system as we have previously described it under the term* gmPOD *in [3].*

3. Space-Time Galerkin Schemes. In this section, we briefly describe how to formulate a general space-time Galerkin approximation to a generic PDE. This regression is then followed by the discussion of low-rank space-time Galerkin schemes on the base of POD reductions of standard Galerkin bases.

Let $\{\hat\psi_1, \dots, \hat\psi_{\hat s}\} \subset H^1(0,T)$ and $\{\hat\nu_1, \dots, \hat\nu_{\hat q}\} \subset H_0^1(\Omega)$ be the POD bases in space and time, respectively. Then, a space-time Galerkin approximation of the generic semilinear equation system

$$\dot v - \Delta v + N(v) = f \quad \text{on } (0,T] \times \Omega, \tag{10a}$$

$$v\big|_{\partial\Omega} = 0 \quad \text{on } (0,T], \tag{10b}$$

$$v\big|_{t=0} = v_0 \quad \text{on } \Omega, \tag{10c}$$

is given as follows:

The approximate solution $\hat v$ is assumed in the space $\hat{\mathcal{S}} \cdot \hat{\mathcal{Y}} := \operatorname{span}\{\hat\psi_j \hat\nu_i\}_{i=1,\dots,\hat q}^{j=1,\dots,\hat s}$. We introduce the formal vectors of the coefficient functions

$$\hat\Upsilon := \begin{bmatrix} \hat\nu_1 \\ \vdots \\ \hat\nu_{\hat q} \end{bmatrix} \quad \text{and} \quad \hat\Psi := \begin{bmatrix} \hat\psi_1 \\ \vdots \\ \hat\psi_{\hat s} \end{bmatrix}$$

and write $\hat v$ as

$$\begin{bmatrix} \hat\psi_1 & \dots & \hat\psi_{\hat q} \end{bmatrix} \otimes \begin{bmatrix} \hat\nu_1 & \dots & \hat\nu_{\hat s} \end{bmatrix} \hat{\mathbf{v}} = [\hat\Psi^{\mathsf{T}} \otimes \hat\Upsilon^{\mathsf{T}}]\hat{\mathbf{v}}, \tag{11}$$

where $\hat{\mathbf{v}} \in \mathbb{R}^{\hat s \hat q}$ is the vector of coefficients. We determine the coefficients by requiring them to satisfy the Galerkin projection of (10a) for every basis function $\hat\nu_i \hat\psi_j$, $i = 1, \dots, \hat q$, $j = 1, \dots, \hat s$

$$\int_0^T \int_\Omega \hat\nu_i \hat\psi_j \dot{\hat v} + \hat\psi_j \nabla\hat\nu_i \nabla\hat v + \hat\nu_i \hat\psi_j N(\hat v) \, \mathrm{d}x \, \mathrm{d}t = \int_0^T \int_\Omega \hat\nu_i \hat\psi_j f \, \mathrm{d}x \, \mathrm{d}t.$$

The latter equations combined give a possibly nonlinear equation system for the vector $\hat{\mathbf{v}}$ of coefficients, which is assembled as follows: For the term with the time derivative we compute

$$\begin{aligned}\int_0^T \int_\Omega [\hat\Psi \otimes \hat\Upsilon] \frac{\partial \hat v}{\partial t} \, \mathrm{d}x \, \mathrm{d}t &= \int_0^T \int_\Omega [\hat\Psi \otimes \hat\Upsilon][\frac{\partial \hat\Psi^\mathsf{T}}{\partial t} \otimes \hat\Upsilon^\mathsf{T}]\hat{\mathbf{v}} \, \mathrm{d}x \, \mathrm{d}t \\ &= \int_0^T \int_\Omega [\hat\Psi \frac{\partial \hat\Psi^\mathsf{T}}{\partial t} \otimes \hat\Upsilon\hat\Upsilon^\mathsf{T}] \, \mathrm{d}x \, \mathrm{d}t\hat{\mathbf{v}} \\ &= [\int_0^T \hat\Psi \frac{\partial \hat\Psi^\mathsf{T}}{\partial t} \, \mathrm{d}t \otimes \int_\Omega \hat\Upsilon\hat\Upsilon^\mathsf{T} \, \mathrm{d}x]\hat{\mathbf{v}} =: [dM_{\hat{\mathcal{S}}} \otimes M_{\hat{\mathcal{Y}}}]\hat{\mathbf{v}}.\end{aligned}$$

By the same principles, for the term with the spatial derivatives, we obtain

$$\begin{aligned}\int_0^T \int_\Omega [\hat\Psi \otimes \nabla\hat\Upsilon]\nabla\hat v \, \mathrm{d}x \, \mathrm{d}t &= \int_0^T \int_\Omega [\hat\Psi \otimes \nabla\hat\Upsilon][\hat\Psi^\mathsf{T} \otimes \nabla\hat\Upsilon^\mathsf{T}]\hat{\mathbf{v}} \, \mathrm{d}x \, \mathrm{d}t \\ &= [\int_0^T \hat\Psi\hat\Psi^T \, \mathrm{d}t \otimes \int_\Omega \nabla\hat\Upsilon\nabla\hat\Upsilon^\mathsf{T} \, \mathrm{d}x]\hat{\mathbf{v}} := [M_{\hat{\mathcal{S}}} \otimes K_{\hat{\mathcal{Y}}}]\hat{\mathbf{v}}.\end{aligned}$$

Note that in higher spatial dimensions, $\nabla\hat v$ as well as $\nabla\hat\nu_i$ is a vector and, thus, in the preceding derivation, $\nabla\hat\Upsilon$ has to be interpreted properly.

Summing up, we can write the overall system as

$$[dM_{\hat{\mathcal{S}}} \otimes M_{\hat{\mathcal{Y}}} + M_{\hat{\mathcal{S}}} \otimes K_{\hat{\mathcal{Y}}}]\hat{\mathbf{v}} + H_{\hat{\mathcal{S}}\hat{\mathcal{Y}}}(\hat{\mathbf{v}}) = f_{\hat{\mathcal{S}}\hat{\mathcal{Y}}}, \tag{12}$$

where

$$M_{\hat{\mathcal{S}}} := [(\hat\nu_i, \hat\nu_j)]_{i,j=1,\dots,\hat s}, \tag{13a}$$

$$dM_{\hat{\mathcal{S}}} := [(\hat\nu_i, \dot{\hat\nu}_j)]_{i,j=1,\dots,\hat s}, \tag{13b}$$

$$M_{\hat{\mathcal{Y}}} := [(\hat\psi_l, \hat\psi_k)]_{l,k=1,\dots,\hat q}, \tag{13c}$$

$$K_{\hat{\mathcal{Y}}} := [(\nabla\hat\psi_l, \nabla\hat\psi_k)]_{l,k=1,\dots,\hat q}, \tag{13d}$$

$$H_{\hat{\mathcal{S}}\hat{\mathcal{Y}}}(\hat{\mathbf{v}}) := [((\hat\nu_i\hat\psi_l, N(\hat v)))]_{i=1,\dots,\hat s;\ l=1,\dots,\hat q}, \tag{13e}$$

and

$$f_{\hat{\mathcal{S}}\hat{\mathcal{Y}}} := [((\hat\nu_i\hat\psi_l, f))]_{i=1,\dots,\hat s;\ l=1,\dots,\hat q}, \tag{13f}$$

are the Galerkin projections of the system operators and the source term assembled in the corresponding inner products.

REMARK 3.1. *In the space-time Galerkin POD context, the reduced bases are projections of standard finite element bases. Concretely, by virtue of Lemma 2.5 and Lemma 2.6 one has that*

$$\hat\Psi = U_{\hat s}^\mathsf{T}\mathbf{L}_{\mathcal{S}}^{-1}\Psi \quad \textit{and} \quad \hat\Upsilon = V_{\hat q}^\mathsf{T}\mathbf{L}_{\mathcal{Y}}^{-1}\Upsilon,$$

where the columns of $U_{\hat s}$ and $V_{\hat q}$ are orthonormal and where $\mathbf{L}_{\mathcal{S}}$ and $\mathbf{L}_{\mathcal{Y}}$ are factors of the mass matrices associated with Ψ and Υ. Accordingly, the coefficients in (13) are *given as*

$$M_{\hat{\mathcal{S}}} := U_{\hat{s}}^{\mathsf{T}} \mathbf{L}_{\mathcal{S}}^{-1} \big[\int_0^T \Psi \Psi^{\mathsf{T}} \, \mathrm{d}s \big] \mathbf{L}_{\mathcal{S}}^{-\mathsf{T}} U_{\hat{s}} = U_{\hat{s}}^{\mathsf{T}} \mathbf{L}_{\mathcal{S}}^{-1} M_{\mathcal{S}} \mathbf{L}_{\mathcal{S}}^{-\mathsf{T}} U_{\hat{s}} = I_{\hat{s}}, \tag{14a}$$

$$dM_{\hat{\mathcal{S}}} := U_{\hat{s}}^{\mathsf{T}} \mathbf{L}_{\mathcal{S}}^{-1} \big[\int_0^T \Psi \dot{\Psi}^{\mathsf{T}} \, \mathrm{d}s \big] \mathbf{L}_{\mathcal{S}}^{-\mathsf{T}} U_{\hat{s}}, \tag{14b}$$

$$M_{\hat{\mathcal{Y}}} := V_{\hat{q}}^{\mathsf{T}} \mathbf{L}_{\mathcal{Y}}^{-1} \big[\int_\Omega \Upsilon \Upsilon^{\mathsf{T}} \, \mathrm{d}x \big] \mathbf{L}_{\mathcal{Y}}^{-\mathsf{T}} V_{\hat{q}} = V_{\hat{q}}^{\mathsf{T}} \mathbf{L}_{\mathcal{Y}}^{-1} M_{\mathcal{Y}} \mathbf{L}_{\mathcal{Y}}^{-\mathsf{T}} V_{\hat{q}} = I_{\hat{q}}, \tag{14c}$$

$$K_{\hat{\mathcal{Y}}} := V_{\hat{q}}^{\mathsf{T}} \mathbf{L}_{\mathcal{Y}}^{-1} \big[\int_\Omega \nabla \Upsilon \nabla \Upsilon^{\mathsf{T}} \, \mathrm{d}x \big] \mathbf{L}_{\mathcal{Y}}^{-\mathsf{T}} V_{\hat{q}}. \tag{14d}$$

Note that, despite their larger size, stiffness matrices of the standard finite element discretization, as they appear in (14b) and (14d), may be assembled much faster than the stiffness matrices $dM_{\hat{\mathcal{S}}}$ and $K_{\hat{\mathcal{Y}}}$ in the formulation given in (13b) and (13d).

4. Implementation Issues. In this section, we adress how to compute the measurement matrices by means of standard tools, how to incorporate the initial and terminal values in the time discretization, and how to preassemble quadratic nonlinearities.

4.1. Computation of the Measurements. We explain how the measurements (cf. Remark 2.1) that are needed for the computation of the optimal low-rank bases (cf. Lemma 2.5 and Lemma 2.6) can be obtained in practical cases.

In the standard *method-of-lines* approach, a $\mathcal{Y}$ will be used as the FE space for a Galerkin spatial discretization that approximates (10a) by an ODE. In a second step, a time integration scheme is employed to approximate the coefficients $v_1, \dots, v_q \colon (0,T] \to \mathbb{R}$ of the solution

$$\bar{v} \colon (0,T] \to \mathcal{Y} \colon t \mapsto \sum_{i=1}^{q} v_i(t) \nu_i$$

of the resulted ODE. With this and with a chosen time measurement space $\mathcal{S}$, a numerical computed measurement in $\mathcal{S} \cdot \mathcal{Y}$ of the actual solution v of (10a), is given as

$$\mathbf{X} = \begin{bmatrix} (v_1, \psi_1)_{\mathcal{S}} & \dots & (v_1, \psi_s)_{\mathcal{S}} \\ \vdots & \ddots & \vdots \\ (v_q, \psi_1)_{\mathcal{S}} & \dots & (v_q, \psi_s)_{\mathcal{S}} \end{bmatrix} \mathbf{M}_{\mathcal{S}}^{-1}. \tag{15}$$

REMARK 4.1. *For smooth trajectories and for measurements using delta distributions centered at some $t_j \in (0,T)$, $j = 1, \dots, s$, with $\int_0^T v_i \delta(t_j) \, \mathrm{d}t = v_i(t_j)$ the matrix (15) degenerates to the standard POD snapshot matrix. In this case, since the delta distributions are not elements of $L^2(0,T)$, there is no way to define an optimal time basis as in Lemma 2.6. However, one can define an optimal low-rank spatial basis by Lemma 2.5 which reduces to the standard POD optimality result with $\mathbf{M}_{\mathcal{S}} = I$, cf. Remarks 2.7 and 2.8.*

4.2. Treatment of the Initial Value. The initial value (10c) requires a special consideration. Firstly, like the solution of the PDE is only well defined when the initial condition is specified, also the space-time Galerkin discretized system (12) is uniquely solvable if an initial condition is provided. Secondly, in particular in view of optimal control, the initial value can be subject to changes which should be realizable in the discretized model.

To maintain the prominent role of the initial condition also in the time discretization, we proceed as follows:

1. We choose an $\mathcal{S}$ that is spanned by a nodal basis $\{\psi_1, \dots, \psi_s\}$ and that ψ_1 is the basis function associated with the node at $t = 0$.
2. For a given function, we compute $\mathbf{X}_0$ as in (4) or (15) setting $\psi_1 = 0$ and $U_{\hat{s},0}$ as the matrix of the $\hat{s} - 1$ leading right singular vectors of $\mathbf{L}_{\mathcal{Y}}^{\mathsf{T}} \mathbf{X}_0 \mathbf{L}_{\mathcal{S}}$.
3. We set

$$U_{\hat{s}} = \begin{bmatrix} \mathbf{L}_{\mathcal{S}}^T \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} & U_{\hat{s},0} \end{bmatrix}$$

and compute the reduced basis as in Lemma 2.6 as

$$\begin{bmatrix} \hat{\psi}_1 \\ \hat{\psi}_2 \\ \vdots \\ \hat{\psi}_{\hat{s}} \end{bmatrix} = U_{\hat{s}}^{\mathsf{T}} \mathbf{L}_{\mathcal{S}}^{-1} \begin{bmatrix} \psi_1 \\ \psi_2 \\ \vdots \\ \psi_s \end{bmatrix}.$$

By this construction we obtain that $\hat{\psi}_1 = \psi_1$ will be associated with the initial value, whereas $\hat{\psi}_2(0) = \ldots = \hat{\psi}_{\hat{s}}(0) = 0$ will still optimally approximate the trajectory.

4.3. Assembling of Quadratic Nonlinearities. As an example, we consider the nonlinearity in the Burgers' equation

$$\frac{1}{2} \partial_x z(t, x)^2 \tag{16}$$

with the spatial coordinate $x \in (0, 1)$, and the time variable $t \in (0, 1]$.

In the time-space Galerkin projection (11), the il-component of the discretized nonlinearity (13e) in the case of (16), is given as

$$\begin{aligned} H_{il}(\hat{\mathbf{v}}) &= \frac{1}{2} \int_0^1 \int_0^1 \hat{\nu}_i \hat{\psi}_l \cdot \partial_x \hat{v}^2 \,\mathrm{d}x \,\mathrm{d}t \\ &= \frac{1}{2} \int_0^1 \int_0^1 \hat{\nu}_i \hat{\psi}_l \cdot \partial_x(([\hat{\Psi}^{\mathsf{T}} \otimes \hat{\Upsilon}^{\mathsf{T}}] \hat{\mathbf{v}})^2) \,\mathrm{d}x \,\mathrm{d}t \\ &= \hat{\mathbf{v}}^{\mathsf{T}} [\int_0^1 \hat{\nu}_i \hat{\Psi} \hat{\Psi}^{\mathsf{T}} \,\mathrm{d}t \otimes \frac{1}{2} \int_0^1 \hat{\psi}_l \partial_x (\hat{\Upsilon} \hat{\Upsilon}^{\mathsf{T}})^2 \,\mathrm{d}x] \hat{\mathbf{v}}, \end{aligned}$$

where we have used the linearity of the *Kronecker product* and that

$$\hat{v}^2 = ([\hat{\Psi}^{\mathsf{T}} \otimes \hat{\Upsilon}^{\mathsf{T}}] \hat{\mathbf{v}})^2 = \hat{\mathbf{v}}^{\mathsf{T}} [\hat{\Psi} \otimes \hat{\Upsilon}] [\hat{\Psi}^{\mathsf{T}} \otimes \hat{\Upsilon}^{\mathsf{T}}] \hat{\mathbf{v}} = \hat{\mathbf{v}}^{\mathsf{T}} [\hat{\Psi} \hat{\Psi}^{\mathsf{T}} \otimes \hat{\Upsilon} \hat{\Upsilon}^{\mathsf{T}}] \hat{\mathbf{v}}.$$

Thus, the evaluation of the discretized nonlinear term can be assisted by precomputing

$$\int_0^1 \hat{\nu}_i \hat{\Psi} \hat{\Psi}^{\mathsf{T}} \,\mathrm{d}t \quad \text{and} \quad \frac{1}{2} \int_0^1 \hat{\psi}_l (\hat{\Upsilon} \partial_x \hat{\Upsilon}^{\mathsf{T}} + \partial_x(\hat{\Upsilon}) \hat{\Upsilon}^{\mathsf{T}}) \,\mathrm{d}x$$

for all $\hat{\nu}_i$, $i = 1, \dots, \hat{s}$ and $\hat{\psi}_l$, $l = 1, \dots, \hat{q}$.

REMARK 4.2. *If $V_{\hat{q}}$ is the matrix of the spatial POD modes that transform the FEM basis Υ into the reduced basis $\hat{\Upsilon}$ via $\hat{\Upsilon} = V_{\hat{q}}^{\mathsf{T}} \mathbf{L}_{\mathcal{Y}}^{-1} \Upsilon$, then the spatial part of the reduced nonlinearity fulfills*

$$\frac{1}{2}\int_0^1 \hat{\psi}_l(\hat{\Upsilon}\partial_x\hat{\Upsilon}^{\mathsf{T}} + \partial_x(\hat{\Upsilon})\hat{\Upsilon}^{\mathsf{T}})\,\mathrm{d}x =$$

$$\frac{1}{2}V_{\hat{q}}^{\mathsf{T}}\mathbf{L}_{\mathcal{Y}}^{-1}\int_0^1 \hat{\psi}_l(\Upsilon\partial_x\Upsilon^{\mathsf{T}} + \partial_x(\Upsilon)\Upsilon^{\mathsf{T}})\,\mathrm{d}x\mathbf{L}_{\mathcal{Y}}^{-\mathsf{T}}V_{\hat{q}},$$

where the inner matrix of the latter expression might be efficiently assembled in a FEM package. The same idea applies to the time-related part.

5. Application in PDE-Constrained Optimization.

We consider a generic optimal control problem.

PROBLEM 5.1. *For a given target trajectory $x^* \in L^2(0,T;L^2(\Omega))$ and a penalization parameter $\alpha > 0$, we consider the optimization problem*

$$\mathcal{J}(x,u) := \frac{1}{2}\|x - x^*\|_{L^2}^2 + \frac{\alpha}{2}\|u\|_{L^2}^2 \to \min_{u\in L^2(0,T;L^2(\Omega))} \tag{17}$$

subject to the generic PDE

$$\dot{x} - \Delta x + N(x) = f + u \quad \text{on } (0,T]\times\Omega, \tag{18a}$$

$$x\big|_{\partial\Omega} = 0 \quad \text{on } (0,T], \tag{18b}$$

$$x\big|_{t=0} = x_0 \quad \text{on } \Omega. \tag{18c}$$

If the nonlinearity is smooth, then necessary optimality conditions with respect to Problem 5.1 for (x,u) are given through $u = \frac{1}{\alpha}\lambda$, where λ solves the adjoint equation

$$-\dot{\lambda} - \Delta\lambda + D_xN(x)^{\mathsf{T}}\lambda + x = x^* \quad \text{on } (0,T]\times\Omega, \tag{19a}$$

$$\lambda\big|_{\partial\Omega} = 0 \quad \text{on } (0,T], \tag{19b}$$

$$\lambda\big|_{t=T} = 0 \quad \text{on } \Omega, \tag{19c}$$

where D_x denotes the Frechét derivative, which is coupled to the state equation (18) through x and u; see [16].

Given low-dimensional spaces $\hat{\mathcal{S}} := \operatorname{span}\{\hat{\psi}_1,\dots,\hat{\psi}_{\hat{s}}\}$, $\hat{\mathcal{R}} := \operatorname{span}\{\hat{\phi}_1,\dots,\hat{\phi}_{\hat{r}}\} \subset H^1(0,T)$ and $\hat{\mathcal{Y}} := \operatorname{span}\{\hat{\nu}_1,\dots,\hat{\nu}_{\hat{q}}\}$, $\hat{\Lambda} := \operatorname{span}\{\lambda_1,\dots,\lambda_{\hat{p}}\} \subset H_0^1(\Omega)$, a tensor space-time Galerkin discretization of the coupled system (18)-(19) reads

$$[dM_{\hat{\mathcal{S}}}\otimes M_{\hat{\mathcal{Y}}} + M_{\hat{\mathcal{S}}}\otimes K_{\hat{\mathcal{Y}}}]\hat{\mathbf{v}} + H_{\hat{\mathcal{S}}\hat{\mathcal{Y}}}(\hat{\mathbf{v}}) - \frac{1}{\alpha}[M_{\hat{\mathcal{S}}\hat{\mathcal{R}}}\otimes M_{\hat{\mathcal{Y}}\hat{\Lambda}}]\hat{\boldsymbol{\lambda}} = f_{\hat{\mathcal{S}}\hat{\mathcal{Y}}}, \tag{20a}$$

$$[-dM_{\hat{\mathcal{R}}}\otimes M_{\hat{\Lambda}} + M_{\hat{\mathcal{R}}}\otimes K_{\hat{\Lambda}}]\hat{\boldsymbol{\lambda}} + D_xN_{\hat{\Lambda}\hat{\mathcal{R}}}^{\mathsf{T}}(\hat{\mathbf{v}})\hat{\boldsymbol{\lambda}} + [M_{\hat{\mathcal{R}}\hat{\mathcal{S}}}\otimes M_{\hat{\Lambda}\hat{\mathcal{Y}}}]\hat{\mathbf{v}} = [M_{\hat{\mathcal{R}}\hat{\mathcal{S}}}\otimes M_{\hat{\Lambda}\hat{\mathcal{Y}}}]\hat{\mathbf{v}}^*, \tag{20b}$$

with the coefficients $dM_{\hat{\mathcal{R}}}$, $M_{\hat{\mathcal{R}}}$, $M_{\hat{\Lambda}}$, $K_{\hat{\Lambda}}$ and the nonlinearity $D_xN_{\hat{\Lambda}\hat{\mathcal{R}}}^{\mathsf{T}}(\hat{\mathbf{v}})\hat{\boldsymbol{\lambda}}$ defined as in (12), with $M_{\hat{\mathcal{S}}\hat{\mathcal{R}}}$, $M_{\hat{\mathcal{R}}\hat{\mathcal{S}}}$, $M_{\hat{\mathcal{Y}}\hat{\Lambda}}$, $M_{\hat{\Lambda}\hat{\mathcal{Y}}}$ denoting the mixed mass matrices like

$$M_{\hat{\mathcal{S}}\hat{\mathcal{R}}} := [(\hat{\psi}_\ell,\hat{\phi}_k)]_{k=1,\dots,\hat{r}}^{\ell=1,\dots,\hat{s}} \in \mathbb{R}^{\hat{s},\hat{r}},$$

with $\hat{\mathbf{v}}^*$ representing the target v^* projected onto $\hat{\mathcal{S}} \cdot \hat{\mathcal{Y}}$, with the spatial boundary conditions resolved in the ansatz spaces, and with accounting for the initial and terminal conditions via requiring

$$\hat{v}(0) = \sum_{j=1}^{\hat{s}} \sum_{i=1}^{\hat{q}} \mathbf{x}^1_{i,j} \hat{\nu}_i \hat{\psi}_j(0) = \Pi_{\hat{\mathcal{Y}}} x_0 \quad \text{and} \quad \hat{\lambda}(T) = \sum_{j=1}^{\hat{r}} \sum_{i=1}^{\hat{p}} \mathbf{x}^1_{i,j} \hat{\mu}_i \hat{\phi}_j(T) = 0,$$

cf. Section 4.2.

6. Numerical Experiments. We consider the optimal control of a Burgers' equation as it was described in [9, 11] and test the proposed space-time Galerkin POD approach. To estimate the performance quantitatively, we run similar tests with a well-established gradient based method.

6.1. Problem and Test Setup. In Problem 5.1, we replace the generic PDE (18) by the one-dimensional Burgers' equation, namely:

$$\dot{x} - \nu \partial_{\xi\xi} x + \frac{1}{2} \partial_\xi (x^2) = u \qquad \text{on } (0, T] \times (0, L), \tag{21a}$$

$$x\big|_{\xi=0, \xi=L} = 0 \qquad \text{on } (0, T], \tag{21b}$$

$$x\big|_{t=0} = x_0 \qquad \text{on } (0, L), \tag{21c}$$

where L and T denote the length of the space and time interval and where $\nu > 0$ is the so-called viscosity paramter. We set $T = 1$ and $L = 1$ and, as the initial value, we take the step function

$$x_0 \colon (0, 1) \to \mathbb{R} \colon \xi \mapsto \begin{cases} 1, & \text{if } \xi \leq 0.5 \\ 0, & \text{if } \xi > 0.5 \end{cases} . \tag{21d}$$

6.1.1. Definition of the Optimal Control Problems. For the first example problem, we define x^* via $x^*(t) = x_0$ as the target. Thus, the concrete optimal control problem which is designed to keep the system in its initial state (cf. Figure 1(c)) reads as follows:

Problem 6.1. *Given parameters ν and α, find $u \in L^2(0, 1; L^2(0, 1))$ such that*

$$\frac{1}{2} \int_0^1 \int_0^1 (x(t, \xi) - x_0)^2 \, \mathrm{d}\xi \, \mathrm{d}t + \frac{\alpha}{2} \int_0^1 \int_0^1 u^2(t, \xi) \, \mathrm{d}\xi \, \mathrm{d}t \to \min_{u \in L^2(0,1;L^2(0,1))} \tag{22}$$

subject to Burgers' equation (21).

As the second testcase, we consider a space-time varying target state. Therefore, we define the function $\chi_\heartsuit \colon (0, 1) \times (0, 1) \to \{0, 1\}$ as the indicator function of a heart-shaped set in the space-time domain as depicted in Figure 2(c).

Problem 6.2. *Given parameters ν and α, find $u \in L^2(0, 1; L^2(0, 1))$ such that*

$$\frac{1}{2} \int_0^1 \int_0^1 (x(t, \xi) - \chi_\heartsuit(t, \xi))^2 \, \mathrm{d}\xi \, \mathrm{d}t + \frac{\alpha}{2} \int_0^1 \int_0^1 u^2(t, \xi) \, \mathrm{d}\xi \, \mathrm{d}t \to \min_{u \in L^2(0,1;L^2(0,1))} \tag{23}$$

subject to Burgers' equation (21).

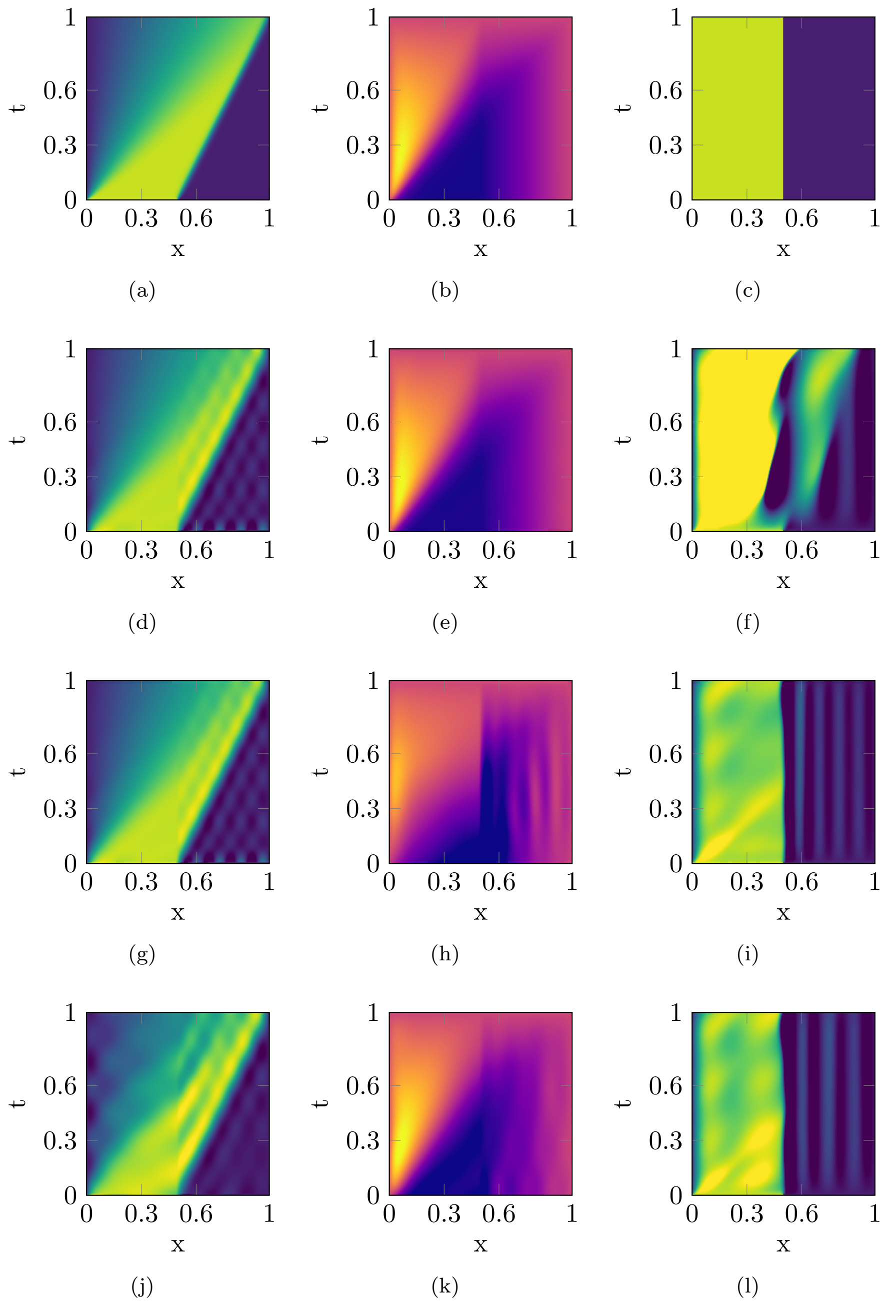

FIG. 1. *Illustration of the effect of the choice of the snapshots on the performance of the low-dimensional approximations of the state (a), the adjoint state (b), and on how the outcome of the optimization matches the target state (c). The second row (d-f) corresponds to the case that snapshots of the state and the adjoint are used to approximate the state and the adjoint, respectively. For the results depicted in the third row (g-i), the optimized basis for the state was used also for the adjoint. The results depicted in the last row (j-l) were obtained by combining state and adjoint snapshots for the computation of the reduced bases. For a comparable illustration we have used color maps with linear intensity on the intervals* $[-0.1, 1.1]$ *for the states and* $[-0.5, 0.5]$ *for the adjoint states. Values that exceeded these margins were cropped.*

6.1.2. Reduced Order Optimization Approaches. We compute suboptimal solutions to the optimal control problems Problem 6.1 and 6.2, i.e, we compute optimal controls on the base of reduced order models that approximate the actual optimal control problems. For that, we consider

- space-time-pod– the space-time Galerkin POD projection of the optimality system; cf. Section 5.
- sqp-pod– a *Sequential-Quadratic Programming* approach for a POD reduced model.

Details on the implementation are given in Sections 6.2 and 6.3 below.

6.1.3. Performance Measures. We measure the performance of the suboptimal controls for the actual optimal control problems Problem 6.1 and 6.2 through:

- The tracking error $\frac{1}{2}\|\hat{x} - x_0\|^2_{L^2}$ or $\frac{1}{2}\|\hat{x} - \chi_{\heartsuit}\|^2_{L^2}$ between the target state and the state $\hat{x}$ achieved by using the suboptimal control $\hat{u}$ in the simulation of the full model.
- The time `walltime` it takes to solve the reduced systems for the suboptimal control $\hat{u}$. We report the lowest measured time out of 5 runs.

Furthermore, we report

- the norm of the computed control $\|u\|^2_{L^2(0,1;L^2(0,1))}$

and, for the tests in the iterative gradient based approach sqp-pod,

- the numbers of function and gradient evaluations `nfc`/`ngc`.

6.1.4. Test Definitions. We investigate the performance of the space-time Galerkin approach over the given range of parameters by means of the following test suites:

- *Dimension of the reduced model* – We gradually increase the degrees of freedom of the reduced model; equally distributed to the space and time dimension.
- *Space vs. time resolution* – Starting from an equal distribution that has proven to perform well, we gradually increase/decrease the dimension of the reduced model in the time dimension while decreasing/increasing its dimension in the space dimension. In other words, we gradually shift the weighting between spatial and time resolution in the reduced model.
- *Performance vs. viscosity* – For a fixed model dimension, we explore the performance versus varying viscosity parameters.
- *Performance vs. regularization* – For a fixed model dimension, we explore the performance versus the regularization parameter α in the cost functional.

It will turn out, that the sqp-pod approach leads to low tracking errors and cost function values but at higher computational costs. As an attempt to reduce the costs, we consider another test suite for the sqp-pod approach:

- *Performance vs. targeted gradient norm* – We gradually increase the value that is the target of the iterative gradient norm minimization in the gradient-based optimization.

References to all results are given in Table 1. More details on the particular test setups and an interpretation of the results are given in Section 6.2 and Section 6.3 for space-time-pod and sqp-pod, respectively. A comparison and an assessment of both methods is given in Section 6.5.

6.1.5. Implementation. The spatial discretization is carried out with the help of the FEM library *FEniCS* [12]. For the time integration, we use *SciPy*'s builtin ODE-integrator `scipy.integrate.odeint`. The norms are approximated in the used

Test Setup	Problem 6.1 – *step-function* space-time-pod	Problem 6.1 – *step-function* sqp-pod	Problem 6.2 – *heart-shape* space-time-pod	Problem 6.2 – *heart-shape* sqp-pod
Dimension of the reduced model	Tab. 4	Tab. 4	Tab. 12	Tab. 12
Space vs. time resolution	Tab. 5	Tab. 5	Tab. 13	Tab. 13
Performance vs. viscosity	Tab. 6, 7	Tab. 6, 7	Tab. 14, 15	Tab. 14, 15
Performance vs. regularization	Tab. 8, 9	Tab. 8, 9	Tab. 16, 17	Tab. 16, 17
Performance vs. gradient norm	—	Tab. 10, 11	—	Tab. 18, 19

TABLE 1
List of numerical experiments and the corresponding tables of results.

FEM space. The implementation and the code for all tests as well as the documentation of the hardware are available from the author's public git repository [8]; see also the section on code availability on page 19.

6.2. Space-time Generalized POD for Optimal Control. The general procedure is as follows:

1. Do at least one forward solve of the state equation (21) and at least one backward solve of the corresponding adjoint equation, cf. (19), to setup generalized measurement matrices of the state and the costate as explained in Section 4.1.
2. Compute optimized space and time bases for the state and the costate as defined in Lemmas 2.5 and 2.6. To account for the initial and the terminal value, one may resort to the procedure explained in Section 4.2.
3. Set up the projected closed-loop optimality system (20) and solve for the optimal costate $\hat{\boldsymbol{\lambda}}$ of the reduced system.
4. Lift $\hat{\boldsymbol{u}} = \frac{1}{\alpha}\hat{\boldsymbol{\lambda}}$ up to the full space-time grid and apply it as suboptimal control to the actual problem.

To solve the nonlinear system (20) for $\hat{\boldsymbol{\lambda}}$, we use *SciPy*'s `scipy.optimize.fsolve` with the associated Jacobian provided as a function which, among others, can be derived from the representation of the nonlinearities as laid out in Section 4.3.

The procedure is defined by several parameters. In the presented examples, we fix $\mathcal{Y} = \Lambda$ and $\mathcal{S} = \mathcal{R}$, corresponding to the initial space and time discretizations, and investigate the influence of the other parameters on the numerical solution of the optimal control problem. See Table 2 for an overview of the parameters and their default values.

Choice of the measurements. The computation of the measurements and the choice of the reduced bases are important parameters of the approach. Generally, the basis of $\hat{\mathcal{S}} \cdot \hat{\mathcal{Y}}$ should be well suited to approximate the state, whereas the basis $\hat{\mathcal{R}} \cdot \hat{\Lambda}$ should well represent the adjoint state. In the optimization case, where the suboptimal input is defined through $\frac{1}{\alpha}\hat{\boldsymbol{\lambda}}$ and its lifting to the full-order space, two other conditions emerge. Firstly, the reduced basis of the adjoint state should also well approximate the optimal control. Secondly, the bases of the state and the adjoint must not be orthogonal or "almost" orthogonal such that the joint mass matrix $[M_{\hat{\mathcal{S}}\hat{\mathcal{R}}} \otimes M_{\hat{\mathcal{Y}}\hat{\Lambda}}]$ degenerates and the contribution of the input in (20a) vanishes.

As illustrated in the plots in Figure 1, the straight-forward approach of constructing the bases for the state by means of state measurements and the basis for the adjoint by means of measurements of the adjoint, well approximates the state and the adjoint but not the coupled problem. It turned out that taking the state measurements to also construct the reduced space for the adjoint gave a better approximation to the

Parameter	Description	Default Values	Range
$\mathcal{Y}$, Λ	Space of piecewise linear finite elements on an equidistant grid of dimension q, p	$q = p = 220$	–
$\mathcal{S}$, $\mathcal{R}$	Space of linear hat functions on an equidistant grid of dimension s, r	$s = r = 120$	–
$\hat{\mathcal{Y}}$, $\hat{\Lambda}$	POD reductions of $\mathcal{Y}$ and Λ of dimension $\hat{q}$, $\hat{p}$; cf. Lemma 2.5	$\hat{q} = \hat{p} = 12$	$6 - 24$
$\hat{\mathcal{S}}$, $\hat{\mathcal{R}}$	POD reductions of $\mathcal{S}$ and $\mathcal{R}$ of dimension $\hat{s}$, $\hat{r}$; cf. Lemma 2.6	$\hat{r} = \hat{s} = 12$	$6 - 24$
α	Regularization parameter in the cost functional (22)	$1 \cdot 10^{-3}$	$2.5 \cdot 10^{-4} - 1.6 \cdot 10^{-2}$
ν	Viscosity parameter in the PDE (21)	$2 \cdot 10^{-3}$	$5 \cdot 10^{-4} - 1.6 \cdot 10^{-2}$

TABLE 2
Description and values of the parameters of the numerical tests with space-time-pod*; cf. Section 6.2*

optimality system while, naturally, only poorly approximating the adjoint. The best results were obtained in combining state and adjoint state measurements to construct the bases.

Thus, for the computation of the optimal bases for the following tests, we combined the measurements obtained from one forward solve with no control and one backward solve with the state from the forward solve and the target state.

In what follows, we report on the performance of space-time-pod in the test setups as defined in Section 6.1.4.

Dimension of the reduced model. We set $\hat{q} = \hat{p} = \hat{r} = \hat{s} = \hat{K}/4$, with $\hat{K} \in \{24, 36, 48, 72, 96\}$. Thus, for every setup, the nonlinear system (20) of dimension $\hat{K}$ has to be solved for the optimal costate $\hat{\boldsymbol{\lambda}}$. The results of these tests are reported in Tables 4 and 12, respectively.

As expected, the larger the reduced model, the lower the achieved values of the cost functional. Also, with growing order of the reduced model, the time needed to solve the corresponding nonlinear system increases drastically.

Space vs. Time Reduction. From the previous tests, we found that in the considered setup, an overall number of $\hat{K} = 48$ modes is a good compromise between accuracy and computation time. In this section, we examine how the distribution of modes between space and time affects the quality of the suboptimal control. Therefore, and for varying increments/decrements j and i, we set $\hat{q} = \hat{p} := 12 \mp j$ and $\hat{s} = \hat{r} := 12 \pm i$. Accordingly, the overall number of degrees of freedom stays more or less the same throughout the tests but we add weight on the approximation of either the time or the space component.

The results are listed in Table 5 and 13. By putting more emphasis on the space component (for Problem 6.1) or on the the time component (for Problem 6.2) it is possible to get a significant increase in the performance. Interestingly, the timings `walltime` vary significantly even for the same overall dimensions of the reduced model. This variance is due to different convergence behavior of the optimization algorithm used to solve the nonlinear system.

Reduced Order Model vs. Viscosity Parameter. In these tests, we examine how the low-rank space-time Galerkin approach performs over a range of viscosity parameters ν.

The results for Problem 6.1 are listed in Tables 7 for $(\hat{q}, \hat{s}) = (\hat{p}, \hat{r}) = (16, 8)$, which was the most beneficial distribution as found in the previous tests, and in Table 6 for $(\hat{q}, \hat{s}) = (\hat{p}, \hat{r}) = (12, 12)$. The optimal distribution $(16, 8)$ has its performance peak at $\nu = 8 \cdot 10^{-3}$ and outperforms the model with the equally distributed modes over almost the whole range.

The results for Problem 6.2 are listed in Tables 15 and 14 for $(\hat{q}, \hat{s}) = (10, 15)$ and for $(\hat{q}, \hat{s}) = (12, 12)$. In terms of the tracking error, the non-equal distribution outperforms the equal distribution over the whole range while in computation time there is no significant difference.

All tests reveal another phenomenon, namely that the control magnitude $\|\hat{u}\|^2_{L^2}$ increases with the parameter ν. This reflects that, in contrast to forward simulations, in control problems, a larger viscosity makes the system harder to solve especially for non-smooth target functions. At the other side of the spectrum, for low values of ν, the problem is *convection dominated* and hard to approximate by POD bases.

Regularization Parameter. In this section, we examine the influence of the regularization parameter α on the performance – for $(\hat{q}, \hat{s}) = (\hat{p}, \hat{r}) \in \{(16, 8), (12, 12)\}$ (for Problem 6.1) and $(\hat{q}, \hat{s}) = (\hat{p}, \hat{r}) \in \{(10, 15), (12, 12)\}$ (for Problem 6.2). The results are reported in Tables 9, 8 and 17, 16, respectively.

With smaller values of α, in all cases, the control magnitude $\|\hat{u}\|^2_{L^2}$ increases. However the tracking error $\frac{1}{2}\|\hat{x} - x_0\|^2_{L^2}$ reaches a minimum and then increases again. A reason for this increase might be that for smaller α, the optimality system (20) is harder to solve for the optimizer.

6.3. Gradient-based Optimal Control with POD. To give also a quantitative estimate of the performance of the space-time POD-reduced Galerkin approach for the solution of PDE constrained optimal control problems, we tackle the same optimization Problem 6.1 with the established approach of SQP [4, 7] with the *BFGS* approximation combined with standard POD for spatial model order reduction.

Shortly spoken, the method of *Sequential Quadratic Programming* (SQP) is an iterative scheme for the minimization of the *reduced cost functional* $\tilde{\mathcal{J}}(u) := \mathcal{J}(x(u), u)$ where the iterant u_{k+1} is given as the argument minimum of the Taylor approximation of $\tilde{\mathcal{J}}(u)$ around u_k truncated after the quadratic term,

$$(24) \quad \tilde{\mathcal{J}}_Q(u; u_k) := \tilde{\mathcal{J}}(u_k) + [\nabla_u \tilde{\mathcal{J}}(u_k)](u - u_k) + \frac{1}{2}(u - u_k)^{\mathsf{T}}[\nabla_{uu}\tilde{\mathcal{J}}(u_k)](u - u_k).$$

We approximate the gradient $\nabla_u \tilde{\mathcal{J}}(u_k)$ and the inverse of the Hessian $\nabla_{uu}\tilde{\mathcal{J}}(u_k)$, as needed for the minimization of $\tilde{\mathcal{J}}_Q(u; u_k)$ in (24), by solving the adjoint equation and employing the BFGS approximation formula [7, Ch. 3.2.1]. Accordingly, in the generic case, the cost of every iteration is basically that of a solve of (18) to obtain the state x_k corresponding to u_k and a solve of (19) to obtain λ_k from $x_k(u_k)$ that defines the current gradient $\nabla_u \tilde{\mathcal{J}}(u_k)$ plus, possibly, another few forward solves to determine the step size of the gradient step (line search).

We realize this iteration for the Burgers' Problem 6.1 with the same parameters as before; cf. Table 2. In particular, the numerical solution of the corresponding PDEs base on the same FEM discretization with $q = 220$ degrees of freedom as in the numerical experiments in Section 6.2 and *SciPy*'s builtin ODE-integrator `scipy.integrate.odeint`.

The SQP approach iterates on fully discrete approximations to the input u_k and

Parameter	Description	Base Value	Range
$\mathcal{Y}$, Λ	Space of piecewise linear finite elements on an equidistant grid of dimension q, p	$q = p = 220$	–
$\mathcal{S}$, $\mathcal{R}$	Space of linear hat functions on an equidistant grid of dimension s, r – to compute the snapshots for the POD	$s = r = 120$	–
$\hat{\mathcal{Y}}$, $\hat{\Lambda}$	POD reductions of $\mathcal{Y}$ and Λ of dimension $\hat{q}$, $\hat{p}$; cf. Lemma 2.5	$\hat{q} = \hat{p} = 18$	$10 - 25$
n_t	dimension of the time grid on which u_k is linearly interpolated	18	$10 - 25$
$\mathtt{tol}_\nabla$	Termination tolerance for the norm of the gradient in the SQP iterations	$2.5 \cdot 10^{-4}$	$1.77 \cdot 10^{-4} - 1 \cdot 10^{-3}$
α	Regularization parameter in the cost functional (22)	$3.125 \cdot 10^{-5}$ ($6.25 \cdot 10^{-5}$)	$7.81 \cdot 10^{-6} - 2.5 \cdot 10^{-4}$
ν	Viscosity parameter in the PDE (21)	$2 \cdot 10^{-3}$	$5 \cdot 10^{-4} - 1.6 \cdot 10^{-2}$

TABLE 3
Description and values of the parameters of the numerical examples of Section 6.3. The value in parentheses refers to Problem 6.2.

the BFGS iteration approximates the full Hessian matrix. Thus, both for efficiency and feasibility, the discrete representation need to be compressed. Therefore, we use the same optimized reduced state spaces $\hat{\mathcal{Y}}$ and $\hat{\Lambda}$ for the forward and adjoint problem, and, accordingly, the spatial dimension of the control as for the space-time Galerkin approach; cf. Table 2. The time dimension of u_k is reduced by considering the linear interpolant on an equidistant time-grid of n_t nodes. Thus, the dimension of the discrete u_k that defines the number of unknowns in the optimization is given as $\hat{q} \cdot n_t$.

As further parameters that influence the performance of the SQP-BFGS iteration, we consider $\mathtt{tol}_\nabla$ – the target tolerance value of the gradient minimization. All approximation defining parameters, as well as the problem parameters ν and α are listed in Table 3.

We use *SciPy*'s routine `scipy.optimize.fmin_bfgs` to solve the reduced discrete optimization problem and then lift the obtained suboptimal control $\hat{\mathbf{u}}$ to the full space and apply it in the unreduced problem.

REMARK 6.3. *In the optimal control of systems, the evaluation of the cost functional $\tilde{J}(u_k)$ and its gradient $\nabla_u \tilde{J}(u_k)$ both base on the same solution $x_k(u_k)$ of the forward problem. In the built-in* Scipy *implementation of the BFGS iteration this redundancy is not considered. To account for that, in the reported `walltime`, we have subtracted the time of the redundant forward solves that we estimate as number of gradient computations times the average time for one forward solve.*

Dimension of the Reduced Model. In this section, we test how the dimension of the reduced model affects the performance. The results are listed in Tables 4 and 12. Generally, for higher model dimension, the tracking error decreases at the expense of higher computation times. For the easier problem Problem 6.1 the tracking error, however, goes up again which might be due to poorer performance of the optimization algorithm.

Reduced Order Model vs. Viscosity. Here, we examine how a reduced model of fixed dimension performs with respect to the viscosity parameter. The results are listed in Tables 6, 7 and 14, 15. For the step-function target (Problem 6.1), the expected behavior can be observed: a performance peak in the middle of the parameter range and a control magnitude that increases with the viscosity (Tables 6, 7). For the harder problem with the heart-shaped target, the performance is good over the whole parameter range without showing the particular patterns except that the computation time increases towards the margins (14, 15).

Regularization Parameter. In Tables 8, 9 and 16, 17, we tabulate the measured performance for the sqp-pod approach versus varying choices of the regularization parameter α as it used in the definition of the cost functionals to penalize the input action. Throughout the investigated range, the performance is equally good. The expected pattern that $\|\hat{u}\|^2_{L^2}$ decreases with increasing α can not be observed.

Target Norm of Gradient. We investigate the influence of the termination criterion for the SQP iteration defined through the target value $\mathtt{tol}_\nabla$ of the gradient $\|\nabla_u \tilde{\mathcal{J}}(u_k)\|$. The results are summarized in Tables 10, 11 and 18, 19. If the threshold value is increased, the computation time decreases at the expense of a higher tracking error. On the other hand, a threshold below a certain value does not have any further effect. This is due to a stagnation in the minimization process and the termination of the iteration because of *precision loss*.

6.4. POD, Space-time POD, and Empirical Interpolation. As illustrated in Section 4.3, for the considered Burgers' equation, the (quadratic) nonlinearity can be reduced in line with the linear terms. However, in more general setups, the question of how to treat a nonlinearity in the reduced equations is immanent.

A generic approach would be an interpolation of $N(\hat{v})$ in the basis of $\hat{\mathcal{S}} \cdot \hat{\mathcal{Y}}$, i.e.

$$N(\hat{v}) \approx \sum_{i=1}^{\hat{q}} \sum_{j=1}^{\hat{s}} n_{ij} \hat{\nu}_i \hat{\psi}_j,$$

for which EIM [1] might be extended to space-time setups.

Also for the SQP-POD approach of Section 6.3, the nonlinearity, which basically is defined through the spatial part of the tensor described in Section 4.3, is preassembled and reduced to the reduced dimension. Thus, the nonlinearity can be efficiently evaluated not resorting to the full dimension.

6.5. Summary and Interpretation of the Numerical Results. In the preceding sections, we have used the proposed space-time Galerkin POD approach (space-time-pod) to compute suboptimal controls for a nonlinear PDE.

As a benchmark, we have solved the same problems with a well-established gradient-based method (sqp-pod). The benchmark implementation is highly optimized in terms of runtime and accuracy. In particular, the space dimension of the forward and backward problem is reduced through POD, the nonlinearities are preassembled for efficient evaluation in the reduced dimension, and the numerical time integration as well as the optimization is done by *SciPy*'s built-in routines.

In the scenario of the time-constant target (Problem 6.1), in terms of the tracking error, the sqp-pod outperformed space-time-pod by a factor of 2. If the optimization in the sqp-pod algorithm is stopped on the tracking error level of space-time-pod, the space-time Galerkin approach appears to be faster by a factor of 4; cf. Tables 11 and

10 for $\texttt{tol}_\nabla = 7.07 \cdot 10^{-4}$ vs. Table 5 for, e.g., $(\hat{q}, \hat{s}) = (16, 8)$. Thus, for this scenario, the sqp-pod approach leads to good controls, while space-time-pod might be of use for the computation of less optimal controls in significantly shorter times.

In the scenario of the heart-shape target (Problem 6.2), the space-time-pod approach reaches the tracking error level of sqp-pod while still being faster by a factor of 5; cf. e.g. the performance tabulation with respect to viscosity – Table 15.

7. Conclusion and Outlook. We have presented a novel approach to low-rank space-time Galerkin approximations that bases on a generalization of classical snapshot-based POD which then can be extended to POD reduction of time discretizations. We have proven optimality of the reduced bases in the relevant function spaces and discussed the numerical implementation.

The space-time Galerkin POD reduction applies well to optimal control problems, as we have illustrated it for the optimal control of a Burgers' equation. Both in terms of computation time for and efficiency of a suboptimal control, the new approach competes well with established gradient-based approaches. In terms of time needed to compute suboptimal controls, the newly proposed approach clearly outperforms the benchmark implementation. In a more challenging setup with a target function varying both with space and time, the proposed space-time Galerkin catches up with the benchmark also in terms of the tracking error.

In the current implementation of the numerical tests, the resulting nonlinear systems were solved by a general purpose routine, namely *MINPACK*'s `HYBRD` [13] as it is included in *Scipy*. It might be worth investigating, whether the performance of the space-time Galerkin approach for optimal control can be improved by better choices and tuning of the optimization routines.

Another possible further improvement and issue to future work concerning the proposed space-time POD in application to optimal control problems lie in the freedom of the choice of the measurement functions [3]. Moreover, the underlying tensor structure is readily extended to include further directions of the state space like parameter dependencies [2] or inputs. Another issue that needs to be addressed is the treatment of general nonlinearities that can not be treated by preassembling like in the presented quadratic case. Then, an inclusion of *empirical interpolation* (EIM) [1] might be needed to achieve efficiency of the reduction. Moreover, it seems worth investigating whether the principles of space-time POD can be used to construct optimized bases for the interpolation.

Code Availability.

The source code of the implementations used to compute the presented results can be obtained from:

`doi:10.5281/zenodo.583296`

and is authored by: Jan Heiland
Please contact Jan Heiland for licensing information

Acknowledgements. We thank Joost van Zwieten, co-developer of *Nutils*[1], for providing benchmarks and valuable insight into space-time discretizations of Burgers' equation.

REFERENCES

[1] Open source finite element toolbox for Python: http://nutils.org

[1] M. Barrault, Y. Maday, N. C. Nguyen, and A. T. Patera. An 'empirical interpolation' method: application to efficient reduced-basis discretization of partial differential equations. *C. R. Math. Acad. Sci. Paris*, 339(9):667–672, 2004.

[2] M. Baumann, P. Benner, and J. Heiland. A generalized POD space-time Galerkin scheme for parameter dependent dynamical systems. Poster at the MoRePaS III - Workshop on "Model Reduction for Parametrized Systems", published in ScienceOpen Posters, 2015. doi:10.14293/P2199-8442.1.SOP-MATH.P8ECXQ.v1.

[3] M. Baumann, J. Heiland, and M. Schmidt. Discrete input/output maps and their relation to Proper Orthogonal Decomposition. In P. Benner, M. Bollhöfer, D. Kressner, C. Mehl, and T. Stykel, editors, *Numerical Algebra, Matrix Theory, Differential-Algebraic Equations and Control Theory*, pages 585–608. Springer International Publishing, 2015.

[4] P. T. Boggs and J. W. Tolle. Sequential quadratic programming for large-scale nonlinear optimization. *J. Comput. Appl. Math.*, 124(1-2):123 – 137, 2000.

[5] K. Carlberg, J. Ray, and B. van Bloemen Waanders. Decreasing the temporal complexity for nonlinear, implicit reduced-order models by forecasting. *Comp. Meth. Appl. Mech. Eng.*, 289:79–103, 2015.

[6] F. Chinesta, A. Ammar, A. Leygue, and R. Keunings. An overview of the Proper Generalized Decomposition with applications in computational rheology. *J. Non-Newtonian Fluid Mech.*, 166(11):578–592, 2011.

[7] P. E. Gill and E. Wong. Sequential quadratic programming methods. In *Mixed integer nonlinear programming*, pages 147–224. Springer, 2012.

[8] J. Heiland. spacetime-genpod-burgers – Python module for space-time-parameter generalized POD for Burgers equation. https://gitlab.mpi-magdeburg.mpg.de/heiland/spacetime-genpod-burgers, 2015.

[9] M. Heinkenschloss. Numerical solution of implicitly constrained optimization problems. Technical Report TR08-05, Department of Computational and Applied Mathematics, Rice University, 2008.

[10] B. N. Khoromskij and C. Schwab. Tensor-structured Galerkin approximation of parametric and stochastic elliptic PDEs. *SIAM J. Sci. Comput.*, 33(1):364–385, 2011.

[11] K. Kunisch and S. Volkwein. Control of the Burgers equation by a reduced-order approach using Proper Orthogonal Decomposition. *J. Optim. Theory Appl.*, 102(2):345–371, 1999.

[12] A. Logg, K. B. Ølgaard, M. E. Rognes, and G. N. Wells. FFC: the FEniCS form compiler. In *Automated Solution of Differential Equations by the Finite Element Method*, pages 227–238. Springer, Berlin, Germany, 2012.

[13] J. More, B. Garbow, and K. Hillstrom. User guide for MINPACK-1. Technical Report ANL-80-74, Argonne National Laboratory, 1980.

[14] S. Murman, L. Diosady, A. Garai, and M. Ceze. A space-time Discontinuous-Galerkin approach for separated flows. Technical report, AIAA Paper 2016-1059, 2016.

[15] C. Schwab and R. Stevenson. Space-time adaptive wavelet methods for parabolic evolution problems. *Math. Comp.*, 78(267):1293–1318, 2009.

[16] F. Tröltzsch. *Optimal control of partial differential equations. Theory, methods and applications.* Providence, RI: American Mathematical Society (AMS), 2010.

[17] J. Van Zwieten, R. A. Henkes, D. R. Van Der Heul, P. I. Rosen Esquivel, B. Sanderse, and C. Vuik. Space-time hp-adaptive DG-FEM schemes for one-dimensional mulitphase flow models. In *10th International Conference on CFD in Oil & Gas, Metallurgical and Process Industries*, pages 491–500, 2014.

[18] S. Volkwein. *Model reduction using proper orthogonal decomposition.* Lecture Notes, Institute of Mathematics and Scientific Computing, University of Graz, 2011.

[19] S. Volkwein and S. Weiland. An algorithm for Galerkin projections in both time and spatial coordinates. *Proc. 17th MTNS*, 2006.

[20] M. Yano, A. T. Patera, and K. Urban. A space-time hp-interpolation-based certified reduced basis method for Burgers' equation. *Math. Models Methods Appl. Sci.*, 24(09):1903–1935, 2014.

Appendix A. Numerical Results for the Step-function Target.

$\hat{K}$	24	36	48	72	96
$\frac{1}{2}\Vert\hat{x}-x_0\Vert^2_{L^2}$	0.0330	0.0280	0.0192	0.0121	0.0104
$\Vert\hat{u}\Vert^2_{L^2}$	4.1598	5.6362	8.3107	11.236	9.6554
walltime $[s]$	0.12	0.49	2.03	36.7	230

$(\hat{q}, n_t)$	$(10,10)$	$(12,12)$	$(15,15)$	$(18,18)$	$(21,21)$	$(25,25)$
$\frac{1}{2}\Vert\hat{x}-x_0\Vert^2_{L^2}$	0.0263	0.0180	0.0121	0.0066	0.0074	0.0092
$\Vert\hat{u}\Vert^2_{L^2}$	9.6605	11.746	21.044	18.948	13.776	9.0060
walltime	4.52	5.86	10.3	21.1	71.6	201
nfc/ngc	113/101	127/115	140/140	175/174	112/112	93/ 93

TABLE 4

Problem 6.1 with space-time-pod *(top) and* sqp-pod *(bottom): Performance of the suboptimal control versus varying resolutions of space and time.*

$(\hat{q},\hat{s})/(\hat{p},\hat{r})$	(16, 8)	(15,10)	(12,10)	(12,12)	(10,12)	(10,15)	(8,16)
$\frac{1}{2}\Vert\hat{x}-x_0\Vert^2_{L^2}$	0.0117	0.0106	0.0151	0.0192	0.0303	0.0318	0.0339
$\Vert\hat{u}\Vert^2_{L^2}$	10.0825	10.631	8.1014	8.3107	7.5060	7.0783	5.0666
walltime	1.58	2.39	1.11	2.09	1.14	1.74	0.98

$(\hat{q}, n_t)$	$(13,18)$	$(15,19)$	$(16,20)$	$(19,15)$	$(20,16)$	$(18,13)$
$\frac{1}{2}\Vert\hat{x}-x_0\Vert^2_{L^2}$	0.0159	0.0100	0.0097	0.0064	0.0061	0.0071
$\Vert\hat{u}\Vert^2_{L^2}$	11.472	14.295	14.068	19.243	18.885	20.781
walltime	7.29	13.2	20	15.7	22.9	11.6
nfc/ngc	109/109	153/145	154/143	161/151	132/132	135/135

TABLE 5

Problem 6.1 with space-time-pod *(top) and* sqp-pod *(bottom): Performance of the suboptimal control versus varying distributions of space and time resolutions; cf. Section 6.2 and Section 6.3.*

ν	$5\cdot10^{-4}$	$1\cdot10^{-3}$	$2\cdot10^{-3}$	$4\cdot10^{-3}$	$8\cdot10^{-3}$	$1.6\cdot10^{-2}$	$3.2\cdot10^{-2}$
$\frac{1}{2}\Vert\hat{x}-x_0\Vert^2_{L^2}$	0.0236	0.0259	0.0263	0.0217	0.0153	0.0123	0.0210
$\Vert\hat{u}\Vert^2_{L^2}$	6.5917	6.8152	7.1356	7.8912	9.6422	10.699	14.028
walltime	2.03	2.08	1.74	1.74	1.59	1.39	1.59

ν	$5\cdot10^{-4}$	$1\cdot10^{-3}$	$2\cdot10^{-3}$	$4\cdot10^{-3}$	$8\cdot10^{-3}$	$1.6\cdot10^{-2}$	$3.2\cdot10^{-2}$
$\frac{1}{2}\Vert\hat{x}-x_0\Vert^2_{L^2}$	0.0169	0.0167	0.0147	0.0087	0.0079	0.0105	0.0138
$\Vert\hat{u}\Vert^2_{L^2}$	9.7995	9.7424	9.2163	16.656	16.363	14.363	19.127
walltime	21.3	16.6	13.3	17	21.8	20.5	17.1
nfc/ngc	203/191	141/141	111/111	134/134	113/113	97/ 97	72/ 72

TABLE 6

Problem 6.1 with space-time-pod *(top) and* sqp-pod *(bottom): Performance of the suboptimal control versus varying diffusion parameters ν for $(\hat{q},\hat{s}) = (\hat{p},\hat{r}) = (12,12)$ for* space-time-pod *and for $(\hat{q}, n_t) = (18,18)$ for* sqp-pod*; cf. Section 6.2 and Section 6.3, respectively.*

ν	$5 \cdot 10^{-4}$	$1 \cdot 10^{-3}$	$2 \cdot 10^{-3}$	$4 \cdot 10^{-3}$	$8 \cdot 10^{-3}$	$1.6 \cdot 10^{-2}$	$3.2 \cdot 10^{-2}$
$\frac{1}{2}\|\hat{x} - x_0\|^2_{L^2}$	0.0246	0.0245	0.0188	0.0126	0.0098	0.0111	0.0198
$\|\hat{u}\|^2_{L^2}$	6.5083	6.7953	7.3451	9.3467	10.281	11.446	13.935
walltime	1.20	2.10	0.90	1.57	1.56	1.49	1.27

ν	$5 \cdot 10^{-4}$	$1 \cdot 10^{-3}$	$2 \cdot 10^{-3}$	$4 \cdot 10^{-3}$	$8 \cdot 10^{-3}$	$1.6 \cdot 10^{-2}$	$3.2 \cdot 10^{-2}$
$\frac{1}{2}\|\hat{x} - x_0\|^2_{L^2}$	0.0174	0.0168	0.0172	0.0100	0.0077	0.0098	0.0112
$\|\hat{u}\|^2_{L^2}$	10.142	9.4359	12.516	18.845	17.036	17.809	27.821
walltime	12.9	11.7	10.3	10.4	15.5	17.3	23
nfc/ngc	166/166	168/160	135/135	127/127	107/107	101/101	109/109

TABLE 7

Problem 6.1 with space-time-pod *(top) and* sqp-pod *(bottom): Performance of the suboptimal control versus varying diffusion parameters ν for $(\hat{q}, \hat{s}) = (\hat{p}, \hat{r}) = (16, 8)$ for* space-time-pod*and for $(\hat{q}, n_t) = (18, 13)$ for* sqp-pod*; cf. Section 6.2 and Section 6.3, respectively.*

α	$2.5 \cdot 10^{-4}$	$5 \cdot 10^{-4}$	$1 \cdot 10^{-3}$	$2 \cdot 10^{-3}$	$4 \cdot 10^{-3}$	$8 \cdot 10^{-3}$	$1.6 \cdot 10^{-2}$
$\frac{1}{2}\|\hat{x} - x_0\|^2_{L^2}$	0.0488	0.0293	0.0192	0.0148	0.0145	0.0168	0.0215
$\|\hat{u}\|^2_{L^2}$	13.046	10.290	8.3107	6.4316	4.7372	3.3371	2.2302
walltime	1.39	1.59	2.03	2.77	3.97	4.51	3.51

α	$3.91 \cdot 10^{-6}$	$7.81 \cdot 10^{-6}$	$1.56 \cdot 10^{-5}$	$3.13 \cdot 10^{-5}$	$6.25 \cdot 10^{-5}$	$1.25 \cdot 10^{-4}$
$\frac{1}{2}\|\hat{x} - x_0\|^2_{L^2}$	0.0067	0.0076	0.0069	0.0066	0.0087	0.0088
$\|\hat{u}\|^2_{L^2}$	18.032	14.408	18.539	18.948	12.454	12.742
walltime	17.6	14.5	21.4	21	14.6	14
nfc/ngc	136/136	113/113	200/189	175/174	136/130	131/122

TABLE 8

Problem 6.1 with space-time-pod *(top) and* sqp-pod *(bottom): Performance of the suboptimal control versus varying regularization parameters for $(\hat{q}, \hat{s}) = (\hat{p}, \hat{r}) = (12, 12)$ for* space-time-pod *and for $(\hat{q}, n_t) = (18, 18)$ for* sqp-pod*; cf. Section 6.2 and Section 6.3, respectively.*

α	$2.5 \cdot 10^{-4}$	$5 \cdot 10^{-4}$	$1 \cdot 10^{-3}$	$2 \cdot 10^{-3}$	$4 \cdot 10^{-3}$	$8 \cdot 10^{-3}$	$1.6 \cdot 10^{-2}$
$\frac{1}{2}\|\hat{x} - x_0\|^2_{L^2}$	0.0137	0.0121	0.0117	0.0124	0.0144	0.0179	0.0237
$\|\hat{u}\|^2_{L^2}$	17.013	13.451	10.082	7.1126	4.8117	3.1661	2.0069
walltime	1.17	1.24	1.60	1.92	0.98	1.61	1.68

α	$3.91 \cdot 10^{-6}$	$7.81 \cdot 10^{-6}$	$1.56 \cdot 10^{-5}$	$3.13 \cdot 10^{-5}$	$6.25 \cdot 10^{-5}$	$1.25 \cdot 10^{-4}$
$\frac{1}{2}\|\hat{x} - x_0\|^2_{L^2}$	0.0070	0.0070	0.0070	0.0071	0.0073	0.0073
$\|\hat{u}\|^2_{L^2}$	20.186	20.302	20.451	20.781	17.063	17.089
walltime	11.4	11.7	11.6	11.6	12.2	12.6
nfc/ngc	133/133	136/136	134/134	135/135	156/149	159/151

TABLE 9

Problem 6.1 with space-time-pod *(top) and* sqp-pod *(bottom): Performance of the suboptimal control versus varying regularization parameters α for for $(\hat{q}, \hat{s}) = (\hat{p}, \hat{r}) = (16, 8)$ for* space-time-pod *and for $(\hat{q}, n_t) = (18, 13)$ for* sqp-pod*; cf. Section 6.2 and Section 6.3, respectively.*

`tol`$_\nabla$	$1.77 \cdot 10^{-4}$	$2.5 \cdot 10^{-4}$	$3.54 \cdot 10^{-4}$	$5 \cdot 10^{-4}$	$7.07 \cdot 10^{-4}$	$1 \cdot 10^{-3}$	$1.41 \cdot 10^{-3}$
$\frac{1}{2}\\|\hat{x} - x_0\\|^2_{L^2}$	0.0069	0.0066	0.0078	0.0142	0.0170	0.0224	0.0247
$\\|\hat{u}\\|^2_{L^2}$	20.045	18.948	14.268	5.9516	4.0778	2.7446	3.4519
`walltime`	21.1	21.1	16.3	7.18	5.83	4.16	3.39
`nfc/ngc`	177/176	175/174	143/142	58/ 58	47/ 47	34/ 34	28/ 28

TABLE 10

Problem 6.1 with sqp-pod*: Performance of the suboptimal control versus varying target values of the gradient minimization* `tol`$_\nabla$ *for* $(\hat{q}, n_t) = (\hat{p}, n_t) = (18, 18)$*; cf. Section 6.3.*

`tol`$_\nabla$	$1.77 \cdot 10^{-4}$	$2.5 \cdot 10^{-4}$	$3.54 \cdot 10^{-4}$	$5 \cdot 10^{-4}$	$7.07 \cdot 10^{-4}$	$1 \cdot 10^{-3}$	$1.41 \cdot 10^{-3}$
$\frac{1}{2}\\|\hat{x} - x_0\\|^2_{L^2}$	0.0071	0.0071	0.0072	0.0078	0.0105	0.0176	0.0186
$\\|\hat{u}\\|^2_{L^2}$	20.922	20.781	17.206	14.402	8.6073	4.3427	4.3608
`walltime`	13.7	11.6	10.2	8.61	5.99	3.28	2.88
`nfc/ngc`	164/152	135/135	117/117	102/102	73/ 73	40/ 40	35/ 35

TABLE 11

Problem 6.1 with sqp-pod*: Performance of the suboptimal control versus varying target values of the gradient minimization* `tol`$_\nabla$ *for* $(\hat{q}, n_t) = (\hat{p}, n_t) = (18, 13)$*.*

Appendix B. Numerical Results for the Heart-shape Target.

$\hat{K}$	24	36	48	72	96
$\frac{1}{2}\|\hat{x}-x_0\|^2_{L^2}$	0.0648	0.0449	0.0357	0.0162	0.0137
$\|\hat{u}\|^2_{L^2}$	4.1339	8.1261	9.5000	12.282	12.578
walltime $[s]$	0.15	0.57	2.50	10.7	63.2

$(\hat{q}, n_t)$	$(10,10)$	$(12,12)$	$(15,15)$	$(18,18)$	$(21,21)$	$(25,25)$
$\frac{1}{2}\|\hat{x}-x_0\|^2_{L^2}$	0.0335	0.0325	0.0294	0.0213	0.0191	0.0176
$\|\hat{u}\|^2_{L^2}$	8.9765	9.2680	10.034	17.073	20.067	21.057
walltime	5.74	7.49	7.60	19.0	88.3	250
nfc/ngc	94/ 84	104/ 93	82/ 70	124/114	165/157	134/125

TABLE 12

Problem 6.2 with space-time-pod (top) and sqp-pod (bottom): Performance of the suboptimal control versus varying resolutions of space and time; cf. Section 6.2 and Section 6.3.

$(\hat{q},\hat{s})/(\hat{p},\hat{r})$	(16, 7)	(15,10)	(12,10)	(12,12)	(10,12)	(10,15)	(7,16)
$\frac{1}{2}\|\hat{x}-x_0\|^2_{L^2}$	0.0512	0.0336	0.0336	0.0357	0.0360	0.0230	0.0301
$\|\hat{u}\|^2_{L^2}$	7.4170	9.1155	8.9908	9.5000	9.2767	9.8665	9.6105
walltime	0.64	3.48	1.64	2.50	1.81	2.45	0.97

$(\hat{q}, n_t)$	$(13,18)$	$(15,19)$	$(16,20)$	$(19,15)$	$(20,16)$	$(18,13)$
$\frac{1}{2}\|\hat{x}-x_0\|^2_{L^2}$	0.0243	0.0261	0.0236	0.0282	0.0259	0.0294
$\|\hat{u}\|^2_{L^2}$	18.968	13.898	17.129	9.2041	12.164	9.0284
walltime	13.2	16.8	25.2	10.7	21.7	8.33
nfc/ngc	108/108	89/ 89	145/139	86/ 74	121/113	84/ 72

TABLE 13

Problem 6.2 with space-time-pod (top) and sqp-pod (bottom): Performance of the suboptimal control versus varying distributions of space and time resolutions; cf. Section 6.2 and Section 6.3.

ν	$5\cdot10^{-4}$	$1\cdot10^{-3}$	$2\cdot10^{-3}$	$4\cdot10^{-3}$	$8\cdot10^{-3}$	$1.6\cdot10^{-2}$	$3.2\cdot10^{-2}$
$\frac{1}{2}\|\hat{x}-x_0\|^2_{L^2}$	0.0383	0.0395	0.0396	0.0372	0.0270	0.1066	0.0440
$\|\hat{u}\|^2_{L^2}$	8.8206	8.4898	8.4899	9.0594	12.613	29.956	30.288
walltime	3.07	3.19	1.71	2.90	2.45	2.06	2.12

ν	$5\cdot10^{-4}$	$1\cdot10^{-3}$	$2\cdot10^{-3}$	$4\cdot10^{-3}$	$8\cdot10^{-3}$	$1.6\cdot10^{-2}$	$3.2\cdot10^{-2}$
$\frac{1}{2}\|\hat{x}-x_0\|^2_{L^2}$	0.0273	0.0311	0.0263	0.0223	0.0261	0.0208	0.0210
$\|\hat{u}\|^2_{L^2}$	20.078	11.560	14.617	20.686	9.8258	17.029	19.335
walltime	42.4	21.2	21.3	22.4	19.7	31.5	35.2
nfc/ngc	256/246	134/122	137/126	144/132	108/ 97	104/ 94	99/ 88

TABLE 14

Problem 6.2 with space-time-pod (top) and sqp-pod (bottom): Performance of the suboptimal control versus varying diffusion parameters ν for $(\hat{q},\hat{s}) = (\hat{p},\hat{r}) = (12,12)$ for space-time-pod and for $(\hat{q}, n_t) = (18,18)$ for sqp-pod; cf. Section 6.2 and Section 6.3.

ν	$5 \cdot 10^{-4}$	$1 \cdot 10^{-3}$	$2 \cdot 10^{-3}$	$4 \cdot 10^{-3}$	$8 \cdot 10^{-3}$	$1.6 \cdot 10^{-2}$	$3.2 \cdot 10^{-2}$
$\frac{1}{2}\|\hat{x} - x_0\|^2_{L^2}$	0.0331	0.0337	0.0316	0.0254	0.0187	0.0538	0.0221
$\|\hat{u}\|^2_{L^2}$	8.9185	8.6870	8.7992	9.4938	10.501	23.865	14.265
`walltime`	2.9	2.34	2.68	2.34	2.62	2.22	2.05

ν	$5 \cdot 10^{-4}$	$1 \cdot 10^{-3}$	$2 \cdot 10^{-3}$	$4 \cdot 10^{-3}$	$8 \cdot 10^{-3}$	$1.6 \cdot 10^{-2}$	$3.2 \cdot 10^{-2}$
$\frac{1}{2}\|\hat{x} - x_0\|^2_{L^2}$	0.0346	0.0294	0.0309	0.0269	0.0223	0.0234	0.0243
$\|\hat{u}\|^2_{L^2}$	11.172	16.089	11.540	14.321	17.264	14.183	12.718
`walltime`	12.9	16.0	10.9	12.9	11.6	18.4	19.3
`nfc/ngc`	125/114	152/140	111/101	131/119	118/106	95/ 86	86/ 78

TABLE 15

Problem 6.2 with space-time-pod *(top) and* sqp-pod *(bottom): for $(\hat{q}, \hat{s}) = (\hat{p}, \hat{r}) = (10, 15)$ for* space-time-pod *and for $(\hat{q}, n_t) = (13, 18)$ for* sqp-pod*; cf. Section 6.2 and Section 6.3.*

α	$2.5 \cdot 10^{-4}$	$5 \cdot 10^{-4}$	$1 \cdot 10^{-3}$	$2 \cdot 10^{-3}$	$4 \cdot 10^{-3}$	$8 \cdot 10^{-3}$	$1.6 \cdot 10^{-2}$
$\frac{1}{2}\|\hat{x} - x_0\|^2_{L^2}$	0.0505	0.0434	0.0357	0.0306	0.0302	0.0346	0.0432
$\|\hat{u}\|^2_{L^2}$	16.376	12.769	9.5000	6.7347	4.6081	3.0341	1.8718
`walltime`	1.83	1.92	2.51	1.11	1.07	1.12	1.16

α	$7.81 \cdot 10^{-6}$	$1.56 \cdot 10^{-5}$	$3.13 \cdot 10^{-5}$	$6.25 \cdot 10^{-5}$	$1.25 \cdot 10^{-4}$	$2.5 \cdot 10^{-4}$
$\frac{1}{2}\|\hat{x} - x_0\|^2_{L^2}$	0.0196	0.0228	0.0229	0.0213	0.0228	0.0281
$\|\hat{u}\|^2_{L^2}$	19.440	13.358	13.688	17.073	13.010	9.0108
`walltime`	22	16.2	18.8	19.1	19.5	14.6
`nfc/ngc`	141/130	109/ 97	132/122	124/114	137/127	110/105

TABLE 16

Problem 6.2 with space-time-pod*: Performance of the suboptimal control versus varying regularization parameters α for $(\hat{q}, \hat{s}) = (\hat{p}, \hat{r}) = (12, 12)$ for* space-time-pod *and for $(\hat{q}, n_t) = (18, 18)$ for* sqp-pod*; cf. Section 6.2 and Section 6.3.*

α	$2.5 \cdot 10^{-4}$	$5 \cdot 10^{-4}$	$1 \cdot 10^{-3}$	$2 \cdot 10^{-3}$	$4 \cdot 10^{-3}$	$8 \cdot 10^{-3}$	$1.6 \cdot 10^{-2}$
$\frac{1}{2}\|\hat{x} - x_0\|^2_{L^2}$	0.0269	0.0245	0.0230	0.0238	0.0274	0.0334	0.0422
$\|\hat{u}\|^2_{L^2}$	21.170	14.176	9.8665	6.8345	4.6528	3.0829	1.9315
`walltime`	1.6	2.11	2.45	3.07	1.18	1.23	1.34

α	$7.81 \cdot 10^{-6}$	$1.56 \cdot 10^{-5}$	$3.13 \cdot 10^{-5}$	$6.25 \cdot 10^{-5}$	$1.25 \cdot 10^{-4}$	$2.5 \cdot 10^{-4}$
$\frac{1}{2}\|\hat{x} - x_0\|^2_{L^2}$	0.0242	0.0242	0.0265	0.0243	0.0266	0.0302
$\|\hat{u}\|^2_{L^2}$	18.339	18.530	13.327	18.968	13.847	9.4158
`walltime`	13.3	15.1	10.3	13.2	13.5	8.8
`nfc/ngc`	130/122	147/138	108/ 98	131/120	138/128	100/ 90

TABLE 17

Problem 6.2 with space-time-pod *(top) and* sqp-pod *(bottom): Performance of the suboptimal control versus varying regularization parameters α for $(\hat{q}, \hat{s}) = (\hat{p}, \hat{r}) = (10, 15)$ for* space-time-pod *and for $(\hat{q}, n_t) = (13, 18)$ for* sqp-pod*; cf. Section 6.2 and Section 6.3.*

$\mathtt{tol}_{\nabla}$	$1.77 \cdot 10^{-4}$	$2.5 \cdot 10^{-4}$	$3.54 \cdot 10^{-4}$	$5 \cdot 10^{-4}$	$7.07 \cdot 10^{-4}$	$1 \cdot 10^{-3}$
$\frac{1}{2}\|\hat{x} - x_0\|^2_{L^2}$	0.0213	0.0213	0.0213	0.0226	0.0280	0.0300
$\|\hat{u}\|^2_{L^2}$	17.073	17.073	17.073	14.149	9.1307	6.9991
walltime	19.1	19.1	19.1	15.4	10.6	9.44
nfc/ngc	124/114	124/114	124/114	93/ 93	66/ 66	58/ 58

TABLE 18

Problem 6.2 with sqp-pod*: Performance of the suboptimal control versus varying target values of the gradient minimization* $\mathtt{tol}_{\nabla}$ *for* $(\hat{q}, n_t) = (\hat{p}, n_t) = (18, 18)$*; cf. Section 6.3.*

$\mathtt{tol}_{\nabla}$	$1.77 \cdot 10^{-4}$	$2.5 \cdot 10^{-4}$	$3.54 \cdot 10^{-4}$	$5 \cdot 10^{-4}$	$7.07 \cdot 10^{-4}$	$1 \cdot 10^{-3}$
$\frac{1}{2}\|\hat{x} - x_0\|^2_{L^2}$	0.0243	0.0243	0.0243	0.0243	0.0302	0.0347
$\|\hat{u}\|^2_{L^2}$	18.968	18.968	18.968	18.968	9.7463	8.0792
walltime	13.1	13.1	13.1	11.2	6.29	4.69
nfc/ngc	131/120	131/120	131/120	108/108	67/ 67	50/ 50

TABLE 19

Problem 6.2 with sqp-pod*: Performance of the suboptimal control versus varying target values of the gradient minimization* $\mathtt{tol}_{\nabla}$ *for* $(\hat{q}, n_t) = (\hat{p}, n_t) = (13, 18)$*.*

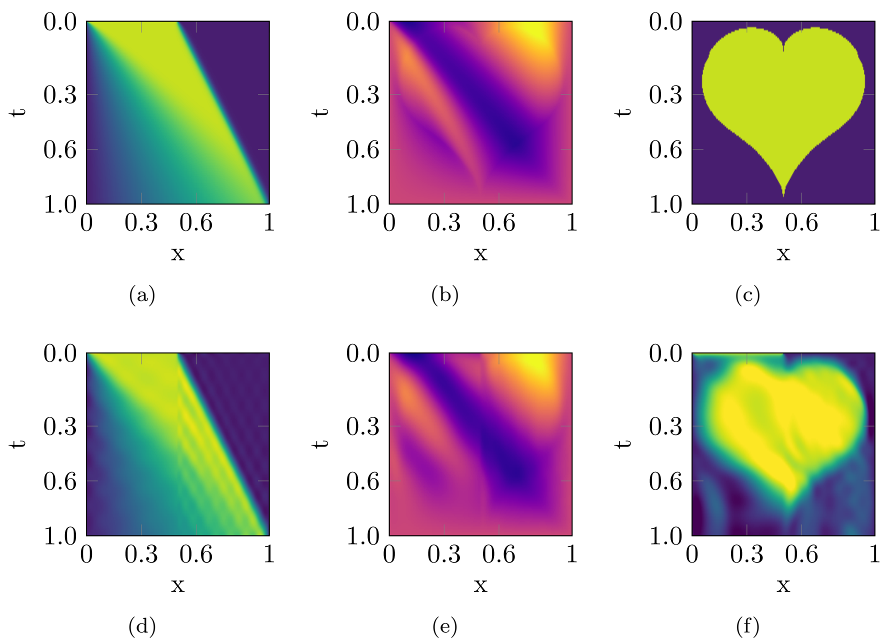

Fig. 2. *Illustration of the optimization problem with a target state varying in space and time as described in Section 6.1.1. The snapshots are taken from a forward simulation without control (a) and the corresponding adjoint solution (b) with respect to the target state (c). Plots (d) and (e) show the results of the low-dimensional space-time Galerkin approximation to the primal and the adjoint state. Plot (f) illustrates the response of a suboptimal control computed on the base of the space-time reduced system. For a comparable illustration we have used color maps with linear intensity on the intervals* $[-0.1, 1.1]$ *for the states and* $[-0.3, 0.3]$ *for the adjoint states. Values that exceeded these margins were cropped.*