MAX PLANCK INSTITUTE
FOR DYNAMICS OF COMPLEX
TECHNICAL SYSTEMS
MAGDEBURG

CSC COMPUTATIONAL METHODS IN
SYSTEMS AND CONTROL THEORY

Space-time Galerkin POD for optimal control of nonlinear PDEs

Manuel Baumann Peter Benner Jan Heiland

September 11, 2018

EUCCO

Outline

Introduction

$\dot{x} - \Delta x = f$

Consider the solution of a PDE:

$$x \in L^2(I; L^2(\Omega))$$

with $I \subset \mathbb{R}$ ... the time-interval
$\Omega \subset \mathbb{R}^n$ ... the spatial domain

and its numerical approximation:

$$\mathbf{x} \in \mathcal{S} \cdot \mathcal{Y}$$

with $\mathcal{S} \subset L^2(I)$ ... discretized time
$\mathcal{Y} \subset L^2(\Omega)$ ... a FE space

Task: Find $\hat{\mathcal{S}} \subset \mathcal{S}$ and $\hat{\mathcal{Y}} \subset \mathcal{Y}$ of much smaller dimension to express $\mathbf{x}$.

Space-Time Spaces

PDE solution $x \in L^2(I; L^2(\Omega))$
$\mathcal{S} \subset L^2(I)$ ... discretized time
$\mathcal{Y} \subset L^2(\Omega)$ ... a FE space

Consider finite dimensional subspaces

$$\mathcal{S} = \operatorname{span}\{\psi_1, \cdots, \psi_s\} \subset L^2(I)$$
$$\mathcal{Y} = \operatorname{span}\{\nu_1, \cdots, \nu_q\} \subset L^2(\Omega)$$

with the mass matrices

$$\mathbf{M}_{\mathcal{S}} = \left[(\psi_i, \psi_j)_{L^2}\right]_{i,j=1,\dots,s} \quad \text{and} \quad \mathbf{M}_{\mathcal{Y}} = \left[(\nu_i, \nu_j)_{L^2}\right]_{i,j=1,\dots,q}$$

and the product space

$$\mathcal{S} \cdot \mathcal{Y} \subset L^2(I; L^2(\Omega)).$$

Space-Time Spaces

We represent a function

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

via its matrix of coefficients

$$\mathbf{X} = \left[\mathbf{x}_{i \cdot j} \right]_{i=1,\dots,q}^{j=1,\dots,s} \in \mathbb{R}^{q,s}$$

and vice versa.

Section 2

Optimal Space Time Product Bases

Space-Time Spaces

Lemma

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

$$\mathbf{X} = \mathbf{M}_{\mathcal{Y}}^{-1} \begin{bmatrix} ((x, \nu_1\psi_1))_{\mathcal{S}\cdot\mathcal{Y}} & \dots & ((x, \nu_1\psi_s))_{\mathcal{S}\cdot\mathcal{Y}} \\ \vdots & \ddots & \vdots \\ ((x, \nu_q\psi_1))_{\mathcal{S}\cdot\mathcal{Y}} & \dots & ((x, \nu_q\psi_s))_{\mathcal{S}\cdot\mathcal{Y}} \end{bmatrix} \mathbf{M}_{\mathcal{S}}^{-1},$$

where

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

Space-Time Spaces

Lemma (Space-time discrete L^2-product)

Let $x^1, x^2 \in \mathcal{S} \cdot \mathcal{Y}$. Then, with

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

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

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

and the induced norm as

$$\|x^\ell\|^2_{\mathcal{S}\cdot\mathcal{Y}} = \|\mathbf{x}^\ell\|^2_{\mathbf{M}_\mathcal{S}\otimes\mathbf{M}_\mathcal{Y}} = \|\mathbf{M}_\mathcal{Y}^{1/2}\mathbf{X}^\ell\mathbf{M}_\mathcal{S}^{1/2}\|^2_F,$$

$\ell = 1, 2.$

Optimal Bases

Lemma (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 subspace $\hat{\mathcal{Y}}$ in the sense that $\|\Pi_{\mathcal{S}\cdot\hat{\mathcal{Y}}} x - x\|_{\mathcal{S}\cdot\mathcal{Y}}$ 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

$$\begin{bmatrix} \hat{\nu}_1 \\ \hat{\nu}_2 \\ \vdots \\ \hat{\nu}_{\hat{q}} \end{bmatrix} = V_{\hat{q}}^{\mathsf{T}} \mathbf{M}_{\mathcal{Y}}^{-1/2} \begin{bmatrix} \nu_1 \\ \nu_2 \\ \vdots \\ \nu_q \end{bmatrix},$$

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

$$\mathbf{M}_{\mathcal{Y}}^{1/2} \mathbf{X} \mathbf{M}_{\mathcal{S}}^{1/2}.$$

Optimal Bases

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

Lemma (Optimal low-rank bases in time[1])

Given $x \in \mathcal{S} \cdot \mathcal{Y}$ and the associated matrix of coefficients $\mathbf{X}$. The best-approximating subspace $\hat{\mathcal{S}}$ in the sense that $\|\Pi_{\hat{\mathcal{S}}\cdot\mathcal{Y}} x - x\|_{\mathcal{S}\cdot\mathcal{Y}}$ 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

$$\begin{bmatrix} \hat{\psi}_1 \\ \hat{\psi}_2 \\ \vdots \\ \hat{\psi}_{\hat{s}} \end{bmatrix} = U_{\hat{s}}^{\mathsf{T}} \mathbf{M}_{\mathcal{S}}^{-1/2} \begin{bmatrix} \psi_1 \\ \psi_2 \\ \vdots \\ \psi_s \end{bmatrix},$$

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

$$\mathbf{M}_{\mathcal{Y}}^{1/2} \mathbf{X} \mathbf{M}_{\mathcal{S}}^{1/2}.$$

[1] see BM&PB&JH '16: *ArXiv:1611.04050*

Section 2

Optimal Space Time Product Bases

Relation to POD

The solution of a spatially discretized PDE

$$x\colon \tau \mapsto \mathbb{R}^q$$

is projected to $\mathcal{S} \cdot \mathbb{R}^q$ via

$$\Pi_{\mathcal{S}\cdot\mathcal{Y}} x = \begin{bmatrix} (x_1, \psi_1)_{L^2} & \dots & (x_1, \psi_s)_{L^2} \\ \vdots & \ddots & \vdots \\ (x_q, \psi_1)_{L^2} & \dots & (x_q, \psi_s)_{L^2} \end{bmatrix} \mathbf{M}_{\mathcal{S}}^{-1}.$$

In the (degenerated) case that ψ_j is a delta distribution centered at $\tau_j \in I$, the coefficient matrix degenerates to

$$\begin{bmatrix} x_1(\tau_1) & \dots & x_1(\tau_s) \\ \vdots & \ddots & \vdots \\ x_q(\tau_1) & \dots & x_q(\tau_s) \end{bmatrix}$$

– the standard POD snapshot matrix.

Section 4

Space-Time Galerkin-POD for Optimal Control

- PDE:

$$\dot{x}(\tau,\xi) + \partial_\xi x(\tau,\xi)^2 = 0 \quad \text{on } I \times \Omega$$

- Ansatz: $x \in \hat{\mathcal{S}} \cdot \hat{\mathcal{Y}}$
  - $\rightarrow$ $x(\tau,\xi) = \sum_{j=1}^{\hat{s}} \sum_{i=1}^{\hat{q}} \mathbf{x}_{i\cdot j}\hat{\psi}_j(\tau)\hat{\nu}_i(\xi)$
  - $\rightarrow$ $x = \begin{bmatrix}\hat{\psi}_1 & \dots & \hat{\psi}_{\hat{q}}\end{bmatrix} \otimes \begin{bmatrix}\hat{\nu}_1 & \dots & \hat{\nu}_{\hat{s}}\end{bmatrix}\mathbf{x} = [\hat{\Psi}^T \otimes \hat{\Upsilon}^T]\mathbf{x}$
  - $\rightarrow$ time derivative: $\dot{x} = [\frac{d}{d\tau}\hat{\Psi}^T \otimes \hat{\Upsilon}^T]\mathbf{x}$

- Space-Time Galerkin Projection:
  - $\rightarrow$ Testfunction $v_{ji} = \hat{\psi}_j\hat{\nu}_i$, $j = 1, \dots, \hat{s}$, $i = 1, \dots, \hat{q}$
  - $\rightarrow$ Galerkin projection

$$\int_{\mathcal{I}}\int_{\Omega}[\hat{\Psi} \otimes \hat{\Upsilon}][\tfrac{d}{d\tau}\hat{\Psi}^T \otimes \hat{\Upsilon}^T]\,\mathrm{d}\tau\,\mathrm{d}\xi\mathbf{x} = -\int_{\mathcal{I}}\int_{\Omega}[\hat{\Psi} \otimes \hat{\Upsilon}]\partial_\xi([\hat{\Psi}^T \otimes \hat{\Upsilon}^T]\mathbf{x})^2\,\mathrm{d}\tau\,\mathrm{d}\xi$$

Tensorization of Quadratic Nonlinearities

With

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

the *ji*-th component of the nonlinearity

$$\begin{aligned}
\int_\mathcal{I}\int_\Omega \hat{\nu}_i\hat{\psi}_j \cdot \partial_\xi \hat{v}^2 \,\mathrm{d}\tau\,\mathrm{d}\xi \\
&= \int_\mathcal{I}\int_\Omega \hat{\nu}_i\hat{\psi}_j \cdot \partial_\xi(([\hat{\Psi}^\mathsf{T} \otimes \hat{\Upsilon}^\mathsf{T}]\hat{\mathbf{x}})^2)\,\mathrm{d}\tau\,\mathrm{d}\xi \\
&= \hat{\mathbf{x}}^\mathsf{T}[\int_I \hat{\nu}_i\hat{\Psi}\hat{\Psi}^\mathsf{T}\,\mathrm{d}\tau \otimes \int_\Omega \hat{\psi}_j\partial_\xi(\hat{\Upsilon}\hat{\Upsilon}^\mathsf{T})^2\,\mathrm{d}\xi]\hat{\mathbf{x}},
\end{aligned}$$

can be efficiently assembled by precomputing

$$\int_I \hat{\nu}_i\hat{\Psi}\hat{\Psi}^\mathsf{T}\,\mathrm{d}\tau \quad \text{and} \quad \int_\Omega \hat{\psi}_j(\hat{\Upsilon}\partial_\xi\hat{\Upsilon}^\mathsf{T} + \partial_\xi(\hat{\Upsilon})\hat{\Upsilon}^\mathsf{T})\,\mathrm{d}\xi.$$

→ Exact hyper-reduction!

→ The reduced model is independent of the full dimensions.

Target: A Space-time Heart Shape

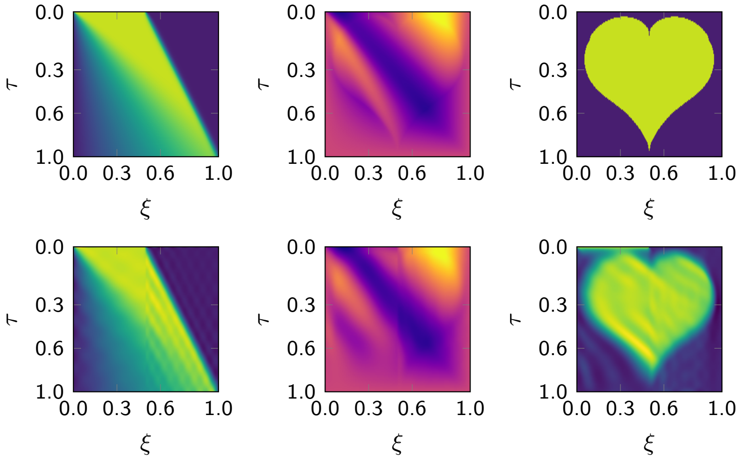


Figure: Illustration of the state, the adjoint, and the target and their approximation via POD-reduced space-time bases.

For a target trajectory $x^* \in L^2(I; L^2(\Omega))$ and a penalization parameter $\alpha > 0$, consider

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

subject to the generic PDE

$$\dot{x} - \Delta x + N(x) = f + u, \quad x(0) = 0. \tag{FWD}$$

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

$$-\dot{\lambda} - \Delta\lambda + D_x N(x)^{\mathsf{T}}\lambda + x = x^*, \quad \lambda(T) = 0. \tag{BWD}$$

Algorithm (space-time-pod):

Offline Phase

1. Do standard forward/backward solves to compute the matrix of measurements for x and λ.
2. Compute optimal low-dimensional spaces $\hat{\mathcal{S}}$, $\hat{\mathcal{R}}$, $\hat{\mathcal{Y}}$, and $\hat{\Lambda}$ for the space and time discretization of the state x and the adjoint state λ.

Online Phase

3. Solve the space-time Galerkin projected necessary optimality conditions (FWD)-(BWD)[2] for the reduced costate $\hat{\boldsymbol{\lambda}}$.

Evaluation

$\rightarrow$ Inflate $\hat{\mathbf{u}} := \frac{1}{\alpha}\hat{\boldsymbol{\lambda}}$ and apply it in the full order simulation.

[2](FWD)-(BWD) is a two-point boundary value problem with initial and terminal conditions for which time stepping schemes like RKM do not apply.

Benchmark: POD and SQP

Algorithm (sqp-pod):

Offline Phase

1. Do standard forward/backward solves to compute the matrix of measurements for x.
2. Compute optimal low-dimensional space $\hat{\mathscr{Y}}$ of dimension $\hat{q}$ via POD.
3. Identify a (manually optimized) time grid of size n_t on which the input is linearly interpolated

→ suboptimal control as minimizer $\hat{\mathbf{u}} \in \mathbb{R}^{\hat{q} \cdot n_t}$ of $\hat{\mathcal{J}}(\mathbf{u}) := \mathcal{J}(x(\mathbf{u}), \mathbf{u})$.

Online Phase

4. Solve $\hat{\mathcal{J}}(\mathbf{u}) \to \min_{\mathbf{u} \in \mathbb{R}^{\hat{q} \cdot n_t}}$ by SQP with BFGS[3] for $\hat{\mathbf{u}} \in \mathbb{R}^{\hat{q} \cdot n_t}$

Evaluation

→ Inflate $\hat{\mathbf{u}}$ and apply it in the full order simulation.

---

[3] here we use MINPACK routines as interfaced in the SciPy optimization module

Numerical Setup

The PDE

- 1D Burger's equation
- $I = (0, 1]$, $\Omega = (0, 1)$
- Viscosity: $\nu = 5 \cdot 10^{-3}$
- Stepfunction as initial value
- Zero Dirichlet conditions

The optimization

- $\alpha = 10^{-3}$ (space-time-pod)
- $\alpha = 6 \cdot 10^{-5}$ (sqp-pod)

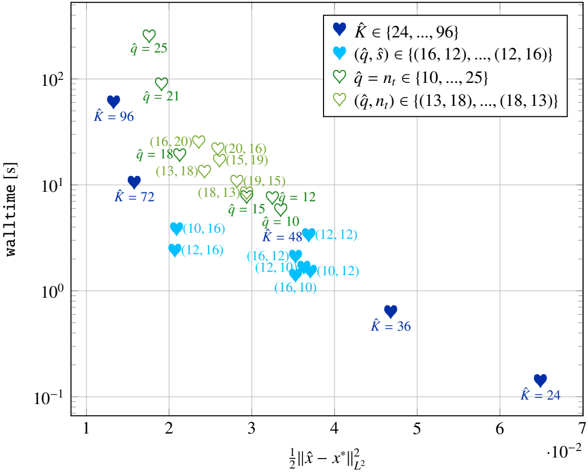
The full model

- Equidistant space and time grids
- $\mathcal{S} = \mathcal{R}$ ... 120 linear hat functions
- $\mathcal{Y} = \Lambda$ ... 220 linear hat functions

The reduced model

- $\hat{\mathcal{Y}} = \hat{\Lambda}$ ... of dimension $\hat{q} = \hat{p}$
- $\hat{\mathcal{S}} \neq \hat{\mathcal{R}}$ ... of dimensions $\hat{s} = \hat{r}$
- $\hat{q}$, $\hat{p}$, $\hat{s}$, $\hat{r}$ ... varying
- n_t ... varying

Performance of the Suboptimal Control



Caption:

The achieved tracking vs. the time needed to compute the suboptimal controls by means of

♡, ♡ ... sqp-pod

♥, ♥ ... space-time-pod.

Parameters:

$\hat{K} :\leftrightarrow \hat{q}, \hat{p}, \hat{r}, \hat{s} = \frac{\hat{K}}{4}$

$(\hat{q}, \hat{s}) = (\hat{p}, \hat{r})$

Conclusion

- The space-time Galerkin POD approach allows for
  - construction of optimized Galerkin bases in space and time
  - in a functional analytical framework
- The resulting space-time Galerkin discretization
  - approximates PDEs by a small system of algebraic equations
  - and naturally extends to boundary value problems in time
  - can be used for efficient computations of (sub)optimal controls
- Future work:
  - Use the functional analytical framework for error estimates.
  - Exploit the freedom of the choice of the measurement functions in $\mathcal{Y}$,
  - to produce, e.g., *optimal* measurements or to compensate for stochastic perturbations.

Further Reading and Coding

M. Baumann, P. Benner, and J. Heiland.
Space-Time Galerkin POD with application in optimal control of semi-linear parabolic partial differential equations.
SIAM J. Sci. Comput.(40). 2018.

M. Baumann, J. Heiland, and M. Schmidt.
Discrete input/output maps and their relation to POD.
In P. Benner et al., editors, *Numerical Algebra, Matrix Theory, Differential-Algebraic Equations and Control Theory*. Springer, 2015.

J. Heiland and M. Baumann.
spacetime-galerkin-pod-bfgs-tests – Python/Matlab implementation space-time POD and BFGS for optimal control of Burgers equation.
2016, `doi:10.5281/zenodo.166339`.

Thank you!

Thank you for your attention!

I am always open for discussion

heiland@mpi-magdeburg.mpg.de
www.janheiland.de
github.com/highlando