Gaussian Process Subspace Regression for Parametric Reduced-Order Modeling

Ruda Zhang[1] Simon Mak[2] David Dunson[1,2]

[1]Department of Mathematics, Duke University
[2]Department of Statistical Science, Duke University

MMLDT-CSET, Sep 26-29, 2021

ruda.zhang@duke.edu
@ZHANG_Ruda

Agenda: Gaussian Process for Reduced Models

- Problem: Parametric Reduced Order Modeling (PROM)
  - approach: adapt local reduced bases.

- Old Idea: Subspace Interpolation
  - polynomial interpolation of subspaces in tangent space
  - ↘ multivalued/discontinuous, deterministic
  - ↘ slow for large-scale systems

- New Idea: Subspace Regression
  - Gaussian process for subspace-valued function
  - ↗ more accurate, smooth, allows UQ
  - ↗ faster, suitable for online computation

RZ, S. Mak, D. Dunson. Gaussian Process Subspace Regression for Model Reduction. arXiv, 2021. https://arxiv.org/abs/2107.04668.

Overview

1. Problem: Parametric Reduced Order Modeling
2. Old Idea: Subspace Interpolation
3. New Idea: Subspace Regression

Reduced Order Modeling (ROM)

- Full model: $\color{red}{\Sigma} = (A, B, C)$, $A \in M_{n,n}$, n is very large.

$$\dot{x} = \boxed{A}\, x + \boxed{B}\, u$$
$$y = \boxed{C}\, x$$

Reduced Order Modeling (ROM)

- Full model: $\Sigma = (A, B, C)$, $A \in M_{n,n}$, n is very large.
- Reduced-order model: $\Sigma_r = (A_r, B_r, C_r)$, $A_r \in M_{k,k}$, k is small.

$$\dot{x} = \boxed{A} \, x + \boxed{B} \, u$$

$$y = \boxed{C} \, x$$

$$\dot{x}_r = \boxed{A_r} \, x_r + \boxed{B_r} \, u$$

$$y_r = \boxed{C_r} \, x_r$$

Reduced Order Modeling (ROM)

- Full model: $\Sigma = (A, B, C)$, $A \in M_{n,n}$, n is very large.
- Reduced-order model: $\Sigma_r = (A_r, B_r, C_r)$, $A_r \in M_{k,k}$, k is small.
- Reduced-order bases V, W for Petrov-Galerkin projection, e.g. by proper orthogonal decomposition (POD).

$$\dot{x} = A x + B u$$
$$y = C x$$

$$\dot{x}_r = A_r x_r + B_r u$$
$$y_r = C_r x_r$$

$$A_r = W^T A V$$
$$B_r = W^T B$$
$$C_r = C V$$

Reduced Order Modeling (ROM)

- Full model: $\Sigma = (A, B, C)$, $A \in M_{n,n}$, n is very large.
- Reduced-order model: $\Sigma_r = (A_r, B_r, C_r)$, $A_r \in M_{k,k}$, k is small.
- Reduced-order bases V, W for Petrov-Galerkin projection, e.g. by proper orthogonal decomposition (POD).
- Parametric full model $\Sigma(\mu), \mu \in \mathcal{P}$; parametric ROM $\Sigma_r(\mu)$.
- Applications: design; optimization; control; uncertainty quantification.

$$\dot{x} = A x + B u$$
$$y = C x$$

$$\dot{x}_r = A_r x_r + B_r u$$
$$y_r = C_r x_r$$

$$A_r = W^T A V$$
$$B_r = W^T B$$
$$C_r = C V$$

Benchmark Problem: Anemometer

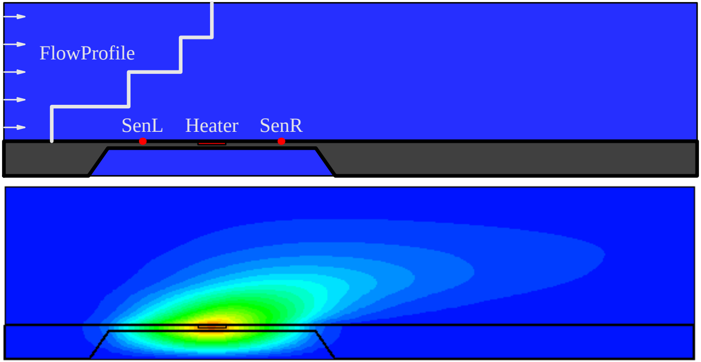
Convection-diffusion PDE:

$$\rho c \partial_t T = \nabla \cdot (\kappa \nabla T) - \rho c v \nabla T + \dot{q}$$

Linearized ODE: ($n = 29{,}008$)

$$E\dot{x} = Ax + bu, \quad y = \langle c, x \rangle$$

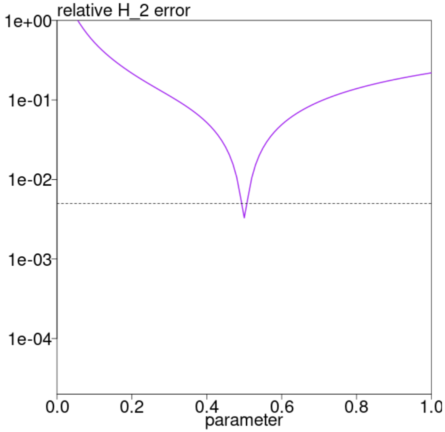
$$A(p) = (1-p)A_1 + pA_2, p \in [0,1]$$



The MORwiki Community. Anemometer. MORwiki – Model Order Reduction Wiki, 2018.

Overview of PROM Methods

- local basis at one point (POD basis, $k = 20$)



Relative $\mathcal{H}_2$ error of a ROM:

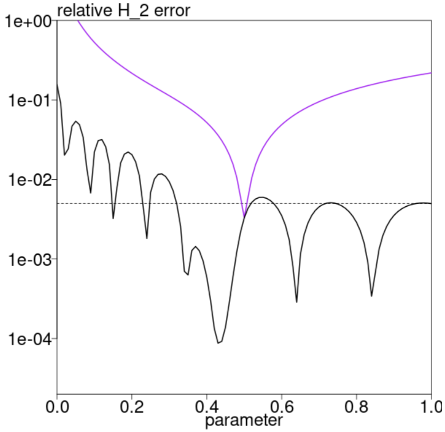
$$\varepsilon = \frac{\|\Sigma - \Sigma_r\|_{\mathcal{H}_2}}{\|\Sigma\|_{\mathcal{H}_2}}$$

$$\|\Sigma - \Sigma_r\|_{\mathcal{H}_2} = \sup_{\mathbf{u} \in \mathcal{L}_2} \frac{\|\mathbf{y} - \mathbf{y}_r\|_{\mathcal{L}_\infty}}{\|\mathbf{u}\|_{\mathcal{L}_2}}$$

P. Benner, S. Gugercin, K. Willcox. A survey of projection-based model reduction methods for parametric dynamical systems. *SIAM Review*, 2015.

Overview of PROM Methods

- local basis at one point (POD basis, $k = 20$)
- local basis at every point



Relative $\mathcal{H}_2$ error of a ROM:

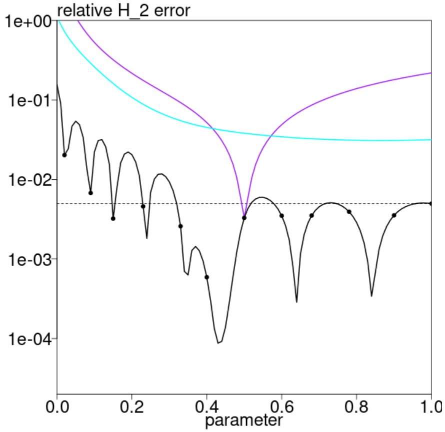
$$\varepsilon = \frac{\|\Sigma - \Sigma_r\|_{\mathcal{H}_2}}{\|\Sigma\|_{\mathcal{H}_2}}$$

$$\|\Sigma - \Sigma_r\|_{\mathcal{H}_2} = \sup_{\mathbf{u}\in\mathcal{L}_2} \frac{\|\mathbf{y} - \mathbf{y}_r\|_{\mathcal{L}_\infty}}{\|\mathbf{u}\|_{\mathcal{L}_2}}$$

P. Benner, S. Gugercin, K. Willcox. A survey of projection-based model reduction methods for parametric dynamical systems. *SIAM Review*, 2015.

Overview of PROM Methods

- local basis at one point (POD basis, $k = 20$)
- local basis at every point
- global basis



Relative $\mathcal{H}_2$ error of a ROM:

$$\varepsilon = \frac{\|\Sigma - \Sigma_r\|_{\mathcal{H}_2}}{\|\Sigma\|_{\mathcal{H}_2}}$$

$$\|\Sigma - \Sigma_r\|_{\mathcal{H}_2} = \sup_{\mathbf{u} \in \mathcal{L}_2} \frac{\|\mathbf{y} - \mathbf{y}_r\|_{\mathcal{L}_\infty}}{\|\mathbf{u}\|_{\mathcal{L}_2}}$$

P. Benner, S. Gugercin, K. Willcox. A survey of projection-based model reduction methods for parametric dynamical systems. *SIAM Review*, 2015.

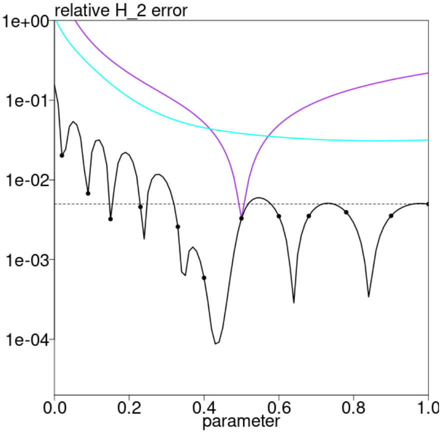
Overview of PROM Methods

- local basis at one point (POD basis, $k = 20$)
- local basis at every point
- global basis
- interpolate local bases*
- interpolate local ROMs
- intrpl. local transfer fn.

Relative $\mathcal{H}_2$ error of a ROM:

$$\varepsilon = \frac{\|\Sigma - \Sigma_r\|_{\mathcal{H}_2}}{\|\Sigma\|_{\mathcal{H}_2}}$$

$$\|\Sigma - \Sigma_r\|_{\mathcal{H}_2} = \sup_{\mathbf{u} \in \mathcal{L}_2} \frac{\|\mathbf{y} - \mathbf{y}_r\|_{\mathcal{L}_\infty}}{\|\mathbf{u}\|_{\mathcal{L}_2}}$$



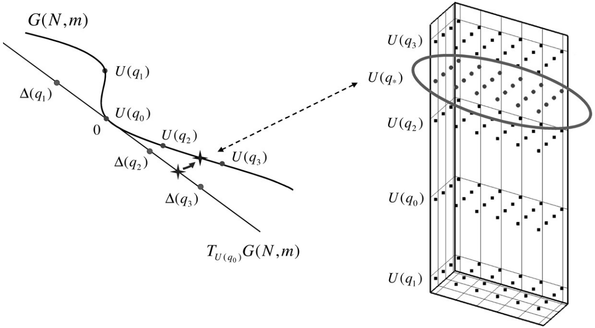
P. Benner, S. Gugercin, K. Willcox. A survey of projection-based model reduction methods for parametric dynamical systems. *SIAM Review*, 2015.

Overview

1. Problem: Parametric Reduced Order Modeling
2. Old Idea: Subspace Interpolation
3. New Idea: Subspace Regression

Subspace Interpolation on the Grassmann Manifold

- ✗ Method: interpolate local bases
- ▶ Reduced bases span the same subspace $\implies$ same ROM
- ▶ Grassmann manifold $G_{k,n} = \{\mathfrak{X} : \mathfrak{X}$ is a k-dim subspace of $\mathbb{R}^n\}$
- ▶ Not a vector space $\implies$ linear combination undefined.
- ✔ Idea: Interpolate tangent vectors.



D. Amsallem, C. Farhat. Interpolation method for adapting ROMs and application to aeroelasticity. *AIAA Journal*, 46(7):1803–1813, July 2008.

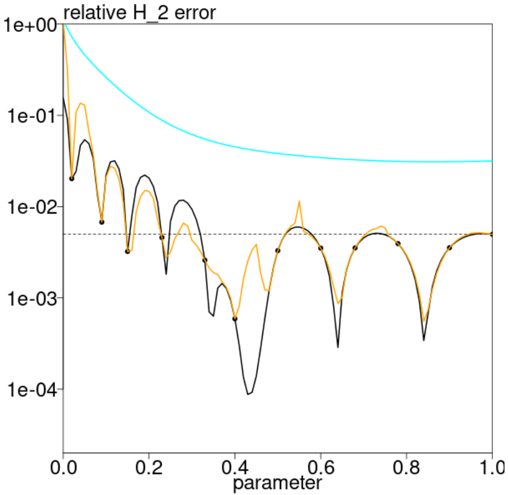
Subspace Interpolation Works (Sometimes)

Setup:

- use local POD basis, $k = 20$
- 12 points
- nearest point as reference

Compare:

- local bases (lower bound)
- global basis (upper bound)
- 5 nearest points in total



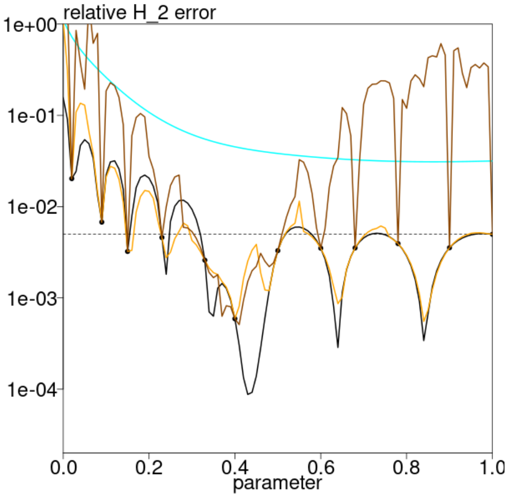
Subspace Interpolation Works (Sometimes)

Setup:

- use local POD basis, $k = 20$
- 12 points
- nearest point as reference

Compare:

- local bases (lower bound)
- global basis (upper bound)
- 5 nearest points in total
- use all 12 points



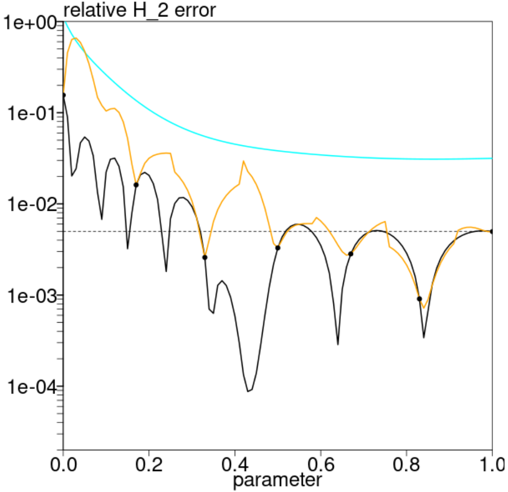
Subspace Interpolation Fails in Small Sample Sizes

Setup:

- use local POD basis, $k = 20$
- 7 points, `p = 0:0.166:1`
- nearest point as reference

Compare:

- local bases (lower bound)
- global basis (upper bound)
- 3 nearest points in total
- Results are similar for $n_r = 4$ or 5.



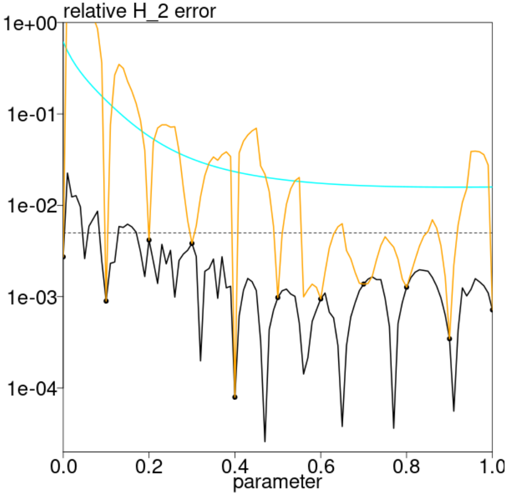
Subspace Interpolation Fails in High Subspace Dimensions

Setup:

- use local POD basis, $k = 40$
- 11 points, `p = 0:0.1:1`
- nearest point as reference

Compare:

- local bases (lower bound)
- global basis (upper bound)
- 5 nearest points in total



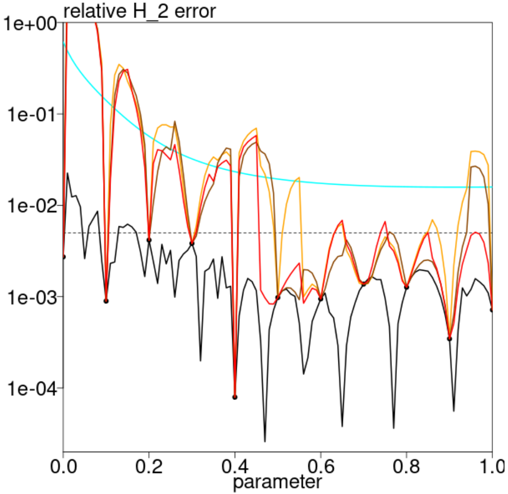
Subspace Interpolation Fails in High Subspace Dimensions

Setup:

- use local POD basis, $k = 40$
- 11 points, `p = 0:0.1:1`
- nearest point as reference

Compare:

- local bases (lower bound)
- global basis (upper bound)
- 5 nearest points in total
- 4 nearest points in total
- 3 nearest points in total



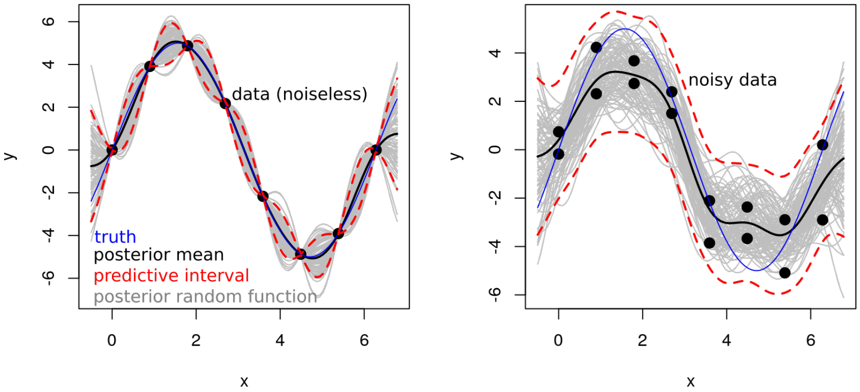
Model selection (ref, n_r, interp. method) is an open problem!

Overview

1. Problem: Parametric Reduced Order Modeling
2. Old Idea: Subspace Interpolation
3. New Idea: Subspace Regression

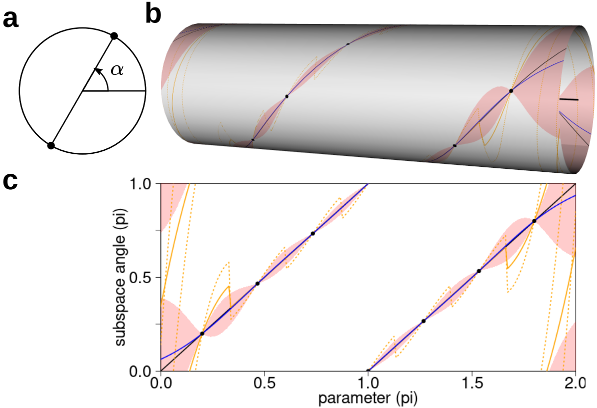
Basics: Gaussian Process (GP) Regression

- Unknown function: $f : \Theta \mapsto \mathbb{R}$, real-valued.
- GP prior process: $f \sim \mathcal{GP}(\mu(x), k(x, x'; \psi))$, hyper-parameters ψ.
- Likelihood: $p(y \mid f, \theta) \sim N(f, \sigma_n^2)$, non-singular Gaussian.
- Conditonal distribution: $p(f_* \mid f, x_*, x)$.
- Posterior distribution: $p(f \mid y, x) \propto p(f \mid x)\ p(y \mid f)$.



Theory: Gaussian Process Subspace (GPS) Model

- task: approximate subspace-valued function $f : \Theta \mapsto G_{k,n}$
- ✗ (classic) GP only works for Euclidean spaces.
- basis representation: $\widetilde{f} : \Theta \mapsto \mathbb{R}^{n \times k}$, $f = \text{span} \circ \widetilde{f}$
- ✔ idea: subspace observation $\implies$ equal likelihood on equivalent bases



The GPS model

Let $\bar{f} : \Theta \mapsto \mathbb{R}^{nk}$ be a representation of f, so that $f = \text{span} \circ \text{vec}^{-1} \circ \bar{f}$, where $\text{vec}^{-1} : \mathbb{R}^{nk} \mapsto M_{n,k}$. We propose a Gaussian process model for $\bar{f}$.

Prior process: $\bar{f} \sim \mathcal{GP}(0, k \otimes \mathbf{I}_{nk})$, with kernel $k : \Theta \times \Theta \mapsto [-1, 1]$. This gives the joint prior:

$$(\mathbf{m}_*, \mathbf{m}) \sim N_{nk(l+1)}(0, \mathbf{K}_{l+1} \otimes \mathbf{I}_{nk}) \tag{1}$$

Assign equal likelihood to equivalence class of representations $[\mathbf{x}_i] = \{\text{vec}(\mathbf{X}_i \mathbf{A}) : \mathbf{A} \in \text{GL}_k\}$:

$$L(\mathbf{m}_i | \mathfrak{X}_i) = 1(\mathbf{m}_i \in [\mathbf{x}_i]) \tag{2}$$

Predictive distribution given observations:

$$\mathbf{m}_* | \mathfrak{X} \sim N_{nk}(0, \mathbf{I}_k \otimes \mathbf{\Sigma}) \tag{3}$$

$$\mathbf{\Sigma}(\boldsymbol{\theta}) = \varepsilon^2 \mathbf{I}_n + \mathbf{X}[\mathbb{X}^T (\widetilde{\mathbf{K}}_l \otimes \mathbf{I}_n) \mathbb{X}]^{-1} \mathbf{X}^T$$

The subspace has a matrix angular central Gaussian distribution $\mathfrak{M}_* | \mathfrak{X} \sim \text{MACG}(\mathbf{\Sigma})$, which degenerates to Uniform$(G_{k,n})$ away from data.

Computational complexity of GPS

Table: Interpolatory PROM methods: flop counts.

	Preprocess	Subspace	ROM	Model selection
GPS	$5nk^2l^2$	k^3l^3	$2k^3l^2$	k^3l^4
Subspace-Int[1]	$10nk^2l^2$	$8nk^2$	$2nk^2$	†
Matrix-Int	$6nk^2l^2$	-	$2k^2l$	†
Manifold-Int[2]	$\mathcal{O}(nk^2l)$	-	$\mathcal{O}(k^3l)^*$	†

* Coefficient usually ~ 50 for matrix exp/log.

† Optimal choice of reference ROM and interpolation scheme is an open problem.

Subspace interpolation is the most used method so far, but it scales with n.
Later developments improve online computation cost, but are not as accurate.

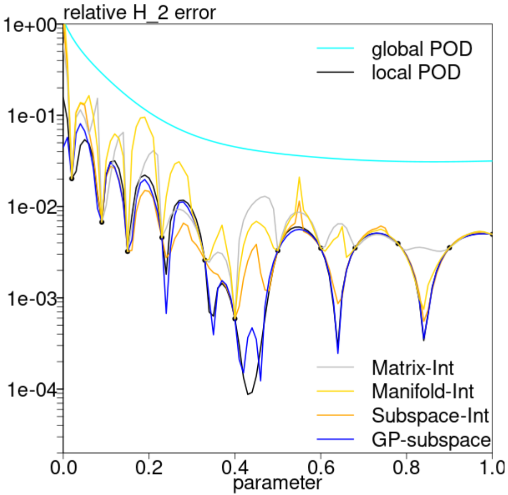
Results: GPS vs. Tangential Interpolation

Setup:

- use local POD basis, $k = 20$
- sample size $l = 12$

Compare:

- local bases (lower bound)
- global basis (upper bound)
- GPS, SE kernel, $\eta = 0.36$
- Subspace interpolation
- Manifold interpolation
- Matrix interpolation

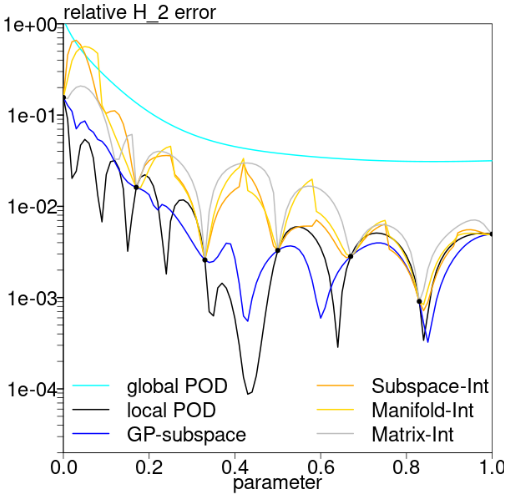


Setup:

- use local POD basis, $k = 20$
- smaller sample size, $l = 7$

Compare:

- local bases (lower bound)
- global basis (upper bound)
- GPS, SE kernel, $\eta = 0.36$
- Subspace interpolation
- Manifold interpolation
- Matrix interpolation



Results: GPS vs. Tangent Interpolation

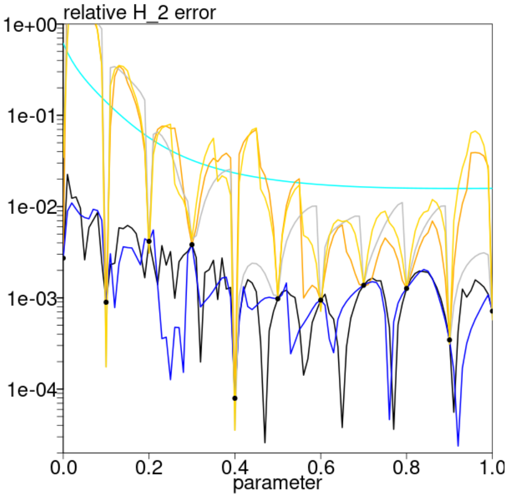
Setup:

- use local POD basis, $k = 40$
- sample size $l = 11$

Compare:

- local bases (lower bound)
- global basis (upper bound)
- GPS, SE kernel, $\eta = 0.25$
- Subspace interpolation
- Manifold interpolation
- Matrix interpolation

GP-subspace retains accuracy of local bases!



Summary

GP subspace regression vs. subspace interpolation:

- ↗ accuracy: data efficient, works for higher dim, allow UQ.
- ↗ speed: independent of system dim, suitable for online computation.

Ongoing and future work:

- local approximate GP for better scalability with many paramters
- adapt covariance matrix for parametric POD (GPSigma, GP-PCA)
- adapt fixed-rank matrix for nonintrusive PROM (GP-SVD)

RZ, S. Mak, D. Dunson. Gaussian Process Subspace Regression for Model Reduction. arXiv, 2021. https://arxiv.org/abs/2107.04668.

Software

R pacakge for GPS: https://github.com/rudazhang/gpsr