UNIVERSITY OF TWENTE.

Model Reduction on Manifolds with Machine Learning

S. Glas[1] P. Buchfink[2] B. Haasdonk[2] B. Unger[2]

[1]University of Twente

[2]University of Stuttgart

Workshop Scientific Machine Learning
CWI, Amsterdam, December 7, 2023

Motivation

- model order reduction (MOR) for dynamical systems
  - given: parametric high-dimensional system of ODEs

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{x}(t;\boldsymbol{\mu}) = \mathbf{f}(\mathbf{x}(t;\boldsymbol{\mu});\boldsymbol{\mu}) \in \mathbb{R}^N, \quad (t;\boldsymbol{\mu}) \in \mathcal{I} \times \mathcal{P} \subset \mathbb{R}^{n_p+1}, \quad N \gg 1$$

  - determine efficient surrogate model for multi-query (optimization, parameter studies, UQ)
  - projection-based MOR

$$\text{full-order model (FOM)} \xrightarrow{\text{projection}} \text{reduced-order model (ROM)}$$

- In this talk:
  1. nonlinear model reduction with MOR on manifolds (Lee, Carlberg, 2020)
  2. structure-preserving algorithms, e.g., symplectic ROM (Peng, Mohseni, 2016), (Maboudi Afkham, Hesthaven, 2017)
  3. where are machine learning methods used in this setting?

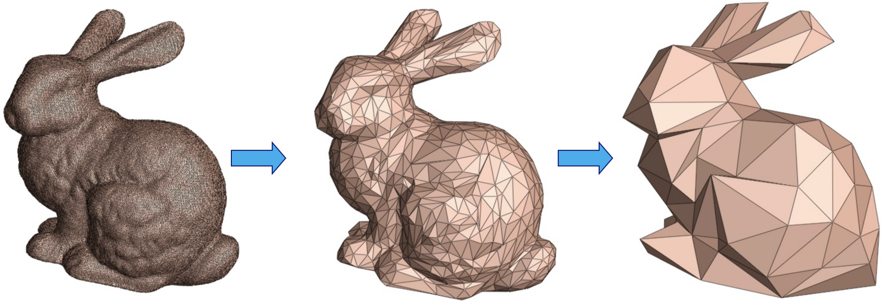
Outlook:

1. Introduction to classical Model Reduction
2. Model Reduction on Manifolds
3. Computation of Reduced Model from Data
4. Conclusion

Outlook:

Aims of Model Reduction

- Reduction of the complexity of the model
- Preservation of approximation quality
- Efficient use of information/avoidance of redundancies
- Error analysis (a priori and a posteriori)



(Petrov–Galerkin) projection-based MOR

high-dimensional ODE system $\quad \frac{\mathrm{d}}{\mathrm{d}t}\mathbf{x}(t;\boldsymbol{\mu}) = \mathbf{f}(\mathbf{x}(t;\boldsymbol{\mu}), t;\boldsymbol{\mu}) \in \mathbb{R}^N$

1. approximation in low-dim subspace of reduced dimension $n \ll N$

$$\mathbf{x}(t;\boldsymbol{\mu}) \approx \mathbf{V}\check{\mathbf{x}}(t;\boldsymbol{\mu}), \qquad \begin{aligned} \check{\mathbf{x}}(t;\boldsymbol{\mu}) \in \mathbb{R}^n &: \text{ reduced state,} \\ \mathbf{V} \in \mathbb{R}^{N\times n} &: \text{ reduced-order basis matrix} \end{aligned}$$

residual: $\mathbf{r}(t;\boldsymbol{\mu}) = \mathbf{V}\frac{\mathrm{d}}{\mathrm{d}t}\check{\mathbf{x}}(t;\boldsymbol{\mu}) - \mathbf{f}(\mathbf{V}\check{\mathbf{x}}(t;\boldsymbol{\mu}), t) \in \mathbb{R}^N$

2. projection of the residual onto a n-dimensional subspace $\mathcal{W} = \mathrm{colspan}(\mathbf{W})$ for $\mathbf{W} \in \mathbb{R}^{N\times n}$

$$\mathbf{W}^\mathsf{T}\mathbf{r}(t;\boldsymbol{\mu}) \stackrel{!}{=} \mathbf{0}_{n\times 1} \Leftrightarrow \underbrace{\mathbf{W}^\mathsf{T}\mathbf{V}\frac{\mathrm{d}}{\mathrm{d}t}\check{\mathbf{x}}(t;\boldsymbol{\mu}) = \mathbf{W}^\mathsf{T}\mathbf{f}(\mathbf{V}\check{\mathbf{x}}, t;\boldsymbol{\mu})}_{\text{reduced system}} \in \mathbb{R}^n$$

When to use linear-subspace MOR ... and when not

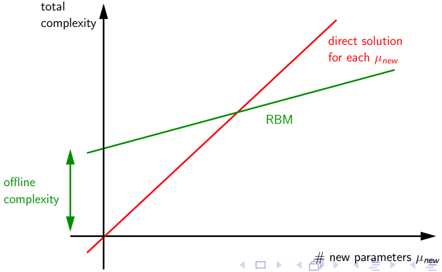
Consequences:

- reduced approximation can only be as good as truth approximation,
- Benchmark: Kolmogorov N-width

$$d_n((\mathcal{P})) := \inf_{V_n;\dim(V_n)=n} \sup_{\boldsymbol{\mu}\in\mathcal{P}} \inf_{\check{\mathbf{v}}\in V_n} \|\mathbf{x}(\boldsymbol{\mu}) - \check{\mathbf{v}}\|$$

- complexity-offset by offline phase $\rightsquigarrow$ MOR (only) reasonable, if

  - many evaluations required ('multi-query context')
  - very fast or limited evaluations required ('realtime context')



Why Model Reduction on Manifolds?

Limitation: Kolmogorov n-width

$$d_n((\mathcal{P})) := \inf_{V_n;\dim(V_n)=n} \sup_{\boldsymbol{\mu}\in\mathcal{P}} \inf_{\check{\mathbf{v}}\in V_n} \|\mathbf{x}(\boldsymbol{\mu}) - \check{\mathbf{v}}\|$$

$\rightsquigarrow$ Slow decay for e.g.,

- Linear transport $d_n(\mathcal{P}) \geq \frac{1}{2} n^{-1/2}$ (Ohlberger, Rave, 2016)
- Wave equations $d_n(\mathcal{P}) \geq \frac{1}{4} n^{-1/2}$ (Greif, Urban, 2019)

Why Model Reduction on Manifolds?

Limitation: Kolmogorov n-width

$$d_n((\mathcal{P})) := \inf_{V_n;\dim(V_n)=n} \sup_{\boldsymbol{\mu}\in\mathcal{P}} \inf_{\check{\mathbf{v}}\in V_n} \|\mathbf{x}(\boldsymbol{\mu}) - \check{\mathbf{v}}\|$$

$\rightsquigarrow$ Slow decay for e.g.,

- Linear transport $d_n(\mathcal{P}) \geq \frac{1}{2} n^{-1/2}$ (Ohlberger, Rave, 2016)
- Wave equations $d_n(\mathcal{P}) \geq \frac{1}{4} n^{-1/2}$ (Greif, Urban, 2019)

How to overcome the Kolmogorov n-width:

1. Update/transform linear subspaces
2. Local linear subspaces
3. Model reduction on nonlinear manifolds (Gu, 2011), (Kashima, 2016), (Hartman, Mestha, 2017), (Lee, Carlberg, 2019), (Lee, Carlberg, 2020), (Kim, Choi, Wideman, Zohdi, 2020), (Fresca, Dedé, Manzoni, 2021), (Rim, Peherstorfer, Mandli, 2023), ...

Outlook:

1. Introduction to classical Model Reduction

2. Model Reduction on Manifolds

3. Computation of Reduced Model from Data

4. Conclusion

FOM and set of all solutions

We assume:

- N-dimensional smooth manifold $\mathcal{M}(=\mathbb{R}^N)$ with $N \gg 1$ large,
- smooth vector field $X(\mu)(=\mathbf{f}(\mathbf{x}(t;\boldsymbol{\mu}),t;\boldsymbol{\mu}) \in \mathbb{R}^N)$
- initial condition $\mathbf{x}_0(\mu) \in \mathcal{M}$, time interval $\mathcal{I} = (t_0, t_\mathrm{f}] \subset \mathbb{R}$

Parametric FOM: find parametric smooth curve $\mathbf{x}(\cdot;\boldsymbol{\mu}) \in \mathcal{C}^\infty(\mathcal{I},\mathcal{M})$ with

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{x}(t;\boldsymbol{\mu}) = \mathbf{f}(\mathbf{x}(t;\boldsymbol{\mu}),t;\boldsymbol{\mu}) \in \mathcal{M}, \quad (t;\boldsymbol{\mu}) \in I \times \mathcal{P} \subset \mathbb{R} \times \mathbb{R}^{n_p}, \quad N \gg 1$$

Goal: approximate set of all solutions

$$S := \left\{\mathbf{x}(t;\boldsymbol{\mu}) \in \mathcal{M} \;\middle|\; (t,\boldsymbol{\mu}) \in \mathcal{I} \times \mathcal{P}\right\} \subset \mathcal{M}$$

Notice: set of all solutions S is typically referred to as solution manifold

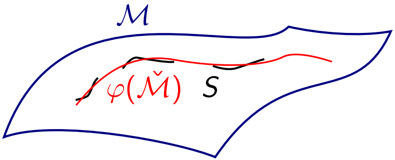
ROM: Embedding

- n-dim smooth manifold $\check{\mathcal{M}}(=\mathbb{R}^n)$, $n \ll N$ and smooth embedding

$$\varphi\colon \check{\mathcal{M}} \to \mathcal{M}$$

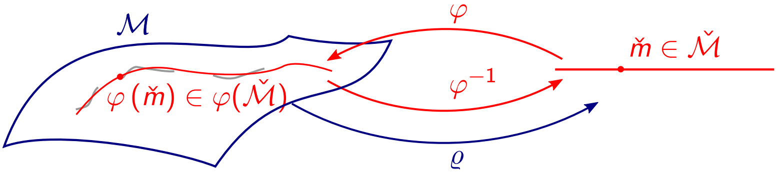
- smooth embedded submanifold $\varphi(\check{\mathcal{M}}) \subset \mathcal{M}$ approximates set of all solutions

$$\varphi(\check{\mathcal{M}}) \approx S$$



ROM: Point Reduction

- mapping points: Find point reduction $\varrho\colon \mathcal{M} \to \check{\mathcal{M}}$



- s.t. the point projection property holds

$$\varrho \circ \varphi = \mathrm{id}_{\check{\mathcal{M}}}$$

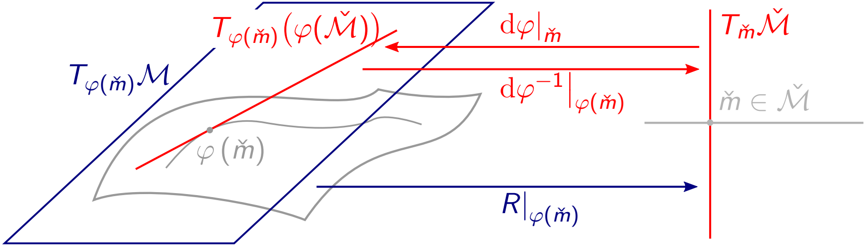
- Then

$$(\varphi \circ \varrho)^2 = \varphi \circ \varrho \circ \varphi \circ \varrho = \varphi \circ \varrho$$

is nonlinear projection

ROM: Tangent Reduction

- mapping tangents at $\check{m} \in \check{\mathcal{M}}$ and $\varphi(\check{m}) \in \varphi(\check{\mathcal{M}})$
- Find tangent reduction $R|_m : T_m\mathcal{M} \to T_{\varrho(m)}\check{\mathcal{M}}$



- s.t. the tangent projection property holds

$$R|_{\varphi(\check{m})} \circ \mathrm{d}\varphi|_{\check{m}} = \mathrm{id}_{T_{\check{m}}\check{\mathcal{M}}} \qquad \text{for all } \check{m} \in \check{\mathcal{M}},$$

Example: Classical Linear-subspace MOR

Projection-based linear-subspace MOR with

- reduced-basis matrix $\mathbf{V} \in \mathbb{R}^{N\times n}$
- projection matrix $\mathbf{W} \in \mathbb{R}^{N\times n}$
- $\mathcal{M} = \mathbb{R}^N$, $\check{\mathcal{M}} = \mathbb{R}^n$
- embedding and the reduction mappings are given by

$$\varrho(\mathbf{m}) := \mathbf{W}^\mathsf{T}\mathbf{m}, \qquad \mathbf{R}|_{\mathbf{m}}(\mathbf{v}) := \mathbf{W}^\mathsf{T}\mathbf{v}, \qquad \varphi(\check{\mathbf{m}}) := \mathbf{V}\check{\mathbf{m}}.$$

- exactly covers case where φ, ϱ linear
- projection properties relate to *biorthogonality* of $\mathbf{W}$ and $\mathbf{V}$

$$\begin{aligned} \varrho \circ \varphi \equiv \mathrm{id}_{\mathbb{R}^n} \quad &\Longleftrightarrow \quad \mathbf{W}^\mathsf{T}\mathbf{V} = \mathbf{I}_n \in \mathbb{R}^{n\times n}, \\ \mathbf{R}|_{\varphi(\check{\mathbf{m}})} \circ \mathrm{d}\varphi|_{\check{\mathbf{m}}} \equiv \mathrm{id}_{\mathbb{R}^n} \quad &\Longleftrightarrow \quad \mathbf{W}^\mathsf{T}\mathbf{V} = \mathbf{I}_n \in \mathbb{R}^{n\times n}, \end{aligned}$$

which is often assumed in linear-subspace MOR.

ROM and State Approximation

Assume given:

- reduced manifold $\check{\mathcal{M}}$ with dimension $n \ll N$
- smooth embedding $\varphi\colon \check{\mathcal{M}} \to \mathcal{M}$
- point reduction $\varrho\colon \mathcal{M} \to \check{\mathcal{M}}$
- tangent reduction $R|_m : T_m\mathcal{M} \to T_{\varrho(m)}\check{\mathcal{M}}$

Parametric ROM: find parametric smooth curve $\check{\mathbf{x}}(\cdot;\mu) \in \mathcal{C}^\infty(\mathcal{I}, \check{\mathcal{M}})$ with

$$\frac{\mathrm{d}}{\mathrm{d}t}\check{\mathbf{x}}(t;\mu) = \mathbf{R}|_{\varphi(\check{\mathbf{x}}(t;\mu))}\mathbf{f}(\varphi(\check{\mathbf{x}}(t;\mu)), t;\mu), \quad \check{\mathbf{x}}(t_0;\mu) = \varrho(\mathbf{x}_0(\mu)) \in \mathbb{R}^n,$$

Reconstruction: approximate the FOM solution curve $\mathbf{x}$ with

$$\mathbf{x}\,(t;\mu) \approx \varphi\,(\check{\mathbf{x}}\,(t;\mu)) \quad \text{for } (t,\mu) \in \mathcal{I} \times \mathcal{P}.$$

Choice of Tangent Reduction: Manifold Petrov–Galerkin (MPG)

Use differential of point reduction map ϱ

$$R|_{\varphi(\check{m})}(v) := \mathrm{d}\varrho|_{\varphi(\check{m})}(v) \in T_{\check{m}}\check{\mathcal{M}}.$$

Then $R|$ is tangent reduction, by differentiating point projection using chain rule

$$\mathrm{d}\varrho|_{\varphi(\check{m})} \circ \mathrm{d}\varphi|_{\check{m}} = \mathrm{id}|_{T_{\check{m}}\check{\mathcal{M}}}$$

Additional structure: Generalized Petrov–Galerkin (GMG)

- additionally given: nondegenerate metrics g and $\check{g}$ such that $\varphi^* g = \check{g}$
- idea: use additional structure to find tangent projection

$$
\begin{tikzcd}
X|_{\varphi(\check{m})} \in & T_{\varphi(\check{m})}\mathcal{M} \arrow[r, "\mathrm{d}\varrho|_{\varphi(\check{m})}"] \arrow[d, "\flat_g"'] & T_{\check{m}}\check{\mathcal{M}} \\
& T^*_{\varphi(\check{m})}\mathcal{M} \arrow[r, "\mathrm{d}\varphi^*|_{\check{m}}"] & T^*_{\check{m}}\check{\mathcal{M}} \arrow[u, "\sharp_{\check{g}}"']
\end{tikzcd}
$$

such that for $\check{m} \in \check{\mathcal{M}}$ and $v \in T_{\varphi(\check{m})}\mathcal{M}$

$$R|_{\varphi(\check{m})}(v) := \left(\sharp_{\check{g}} \circ \mathrm{d}\varphi^*|_{\check{m}} \circ \flat_g\right)(v)\,.$$

- GMG used for Lagrangian systems (LMG) and Hamiltonian systems (SMG)

Outlook:

1. Introduction to classical Model Reduction
2. Model Reduction on Manifolds
3. $\boxed{\text{Computation of Reduced Model from Data}}$
4. Conclusion

Nonlinear Compressive Approximation

Here, embedding and point reduction are defined by

$$\varphi\left(\check{\mathbf{m}}\right) := \mathbf{A}_2\mathbf{f}\left(\check{\mathbf{m}}\right) + \mathbf{A}_1\check{\mathbf{m}} + \mathbf{A}_0, \qquad \varrho\left(\mathbf{m}\right) := \mathbf{B}^{\mathsf{T}}\left(\mathbf{m} - \mathbf{A}_0\right), \tag{1}$$

where

- $\mathbf{A}_2 \in \mathbb{R}^{N\times\widetilde{n}}$, $\mathbf{A}_1 \in \mathbb{R}^{N\times n}$, $\mathbf{A}_0 \in \mathbb{R}^{N}$, $\mathbf{B} \in \mathbb{R}^{N\times n}$
- $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^{\widetilde{n}}$ is a nonlinear mapping for a given $\widetilde{n} \in \mathbb{N}$.

Under the following conditions a point projection property can be shown

$$\mathbf{B}^{\mathsf{T}}\mathbf{A}_1 = \mathbf{I}_n,\ \mathbf{B}^{\mathsf{T}}\mathbf{A}_2 = \mathbf{0}_{n\times\widetilde{n}}$$

Known works include:

- quadratic embeddings, when choosing $\mathbf{f}\left(\check{\mathbf{m}}\right) = \check{\mathbf{m}}^{\otimes^{s}2}$. (Geelen, Wright, Wilcox, 2022)
- Via artificial neural networks (Barnett, Farhat, Maday, 2023)
- NCA has its limitations in terms of the Kolmogorov-($\widetilde{n} + n$)-width (Buchfink, G. Haasdonk, 2023)

Point Projection Property for Autoencoder

Consider ϱ as *encoder*, φ as the *decoder*, and the concatenation $\varphi \circ \varrho$ as an *autoencoder*

- finite subset $S_{\text{train}} \subseteq S$ of set of all solutions $S \subseteq \mathcal{M}$
- its elements $m_{\text{train}} \in S_{\text{train}}$ as *snapshots*
- loss function

$$\mathbf{L}\left(\varphi, \varrho\right) = \frac{1}{|S_{\text{train}}|} \sum_{\mathbf{m}_{\text{train}} \in S_{\text{train}}} \left\| \mathbf{m}_{\text{train}} - \left(\varphi \circ \varrho\right)\left(\mathbf{m}_{\text{train}}\right) \right\|^2.$$

Theorem (Buchfink, G. Haasdonk, Unger 2023)

Assume

- *φ and ϱ are Lipschitz continuous with $C_\varphi, C_\varrho \geq 0$*

Then, the point projection properties are fulfilled approximately s.t. for each $\check{\mathbf{m}} \in \mathbb{R}^n$

$$\begin{aligned} \left\| \left(\varrho \circ \varphi\right)\left(\check{\mathbf{m}}\right) - \check{\mathbf{m}} \right\| \leq\ & C_\varrho \sqrt{|S_{\text{train}}| L\left(\varphi, \varrho\right)} \\ & + \left(C_\varrho C_\varphi + 1\right) \min_{\mathbf{w}_{\text{train}} \in S_{\text{train}}} \left\| \check{\mathbf{m}} - \varrho\left(\mathbf{w}_{\text{train}}\right) \right\| \end{aligned}$$

Special case of GMG with Autoencoder:

- Manifold Galerkin (Lee, Carlberg, 2020) for Riemannian metric $\mathbf{g} \equiv \mathrm{const}$ in coordinates

$$(\mathbf{R}(\mathbf{X}))|_{\check{\mathbf{m}}} = \underbrace{\left(\mathbf{D}\varphi|_{\check{\mathbf{m}}}^{\mathsf{T}} \mathbf{g} \mathbf{D}\varphi|_{\check{\mathbf{m}}}\right)^{-1} \mathbf{D}\varphi|_{\check{\mathbf{m}}}^{\mathsf{T}} \mathbf{g} \mathbf{X}}_{=\left(\mathbf{g}^{1/2}\mathbf{D}\varphi|_{\check{\mathbf{m}}}\right)^{\dagger} \mathbf{g}^{1/2}\mathbf{X}|_{\varphi(\check{\mathbf{m}})}}|_{\varphi(\check{\mathbf{m}})} \in \mathbb{R}^{n}$$

  using the Moore–Penrose pseudoinverse $(\cdot)^{\dagger}$

- Hamiltonian systems, symplectic manifold Galerkin (Buchfink, G. Haasdonk, 2023): choose non-degenerate metric as canonical symplectic form $\mathbf{g} = \mathbb{J}_{2n}^{\mathsf{T}}$

$$(\mathbf{R}(\mathbf{X}))|_{\check{\mathbf{m}}} = \underbrace{\left(\mathbf{D}\varphi|_{\check{\mathbf{m}}}^{\mathsf{T}} \mathbb{J}_{2N} \mathbf{D}\varphi|_{\check{\mathbf{m}}}\right)}_{=\mathbb{J}_{2n}^{\mathsf{T}}}{}^{-1} \mathbf{D}\varphi|_{\check{\mathbf{m}}}^{\mathsf{T}} \mathbb{J}_{2N}^{\mathsf{T}} \mathbf{X}|_{\varphi(\check{\mathbf{m}})} \in \mathbb{R}^{n}.$$

Example: (Parametric) Wave Equation

Linear 1D Wave Equation: $\Omega := [-0.5, 0.5]$, $t \in [0,1]$, $\mu \in [5/12, 5/6]$

$$\begin{aligned} \partial_{tt}^2 u(t,\xi;\mu) &= \mu^2 \partial_{\xi\xi}^2 u(t,\xi;\mu), && \text{on } I \times \Omega, \\ u(0,\xi;\mu) &= u_0(\xi;\mu), && \forall\, \xi \in \Omega, \\ \partial_t u(0,\xi;\mu) &= -\mu\, \partial_\xi u_0(\xi;\mu), && \forall\, \xi \in \Omega, \\ u(t,\xi;\mu) &= 0, && \forall\, t \in I, \xi \in \{-1/2, 1/2\}, \end{aligned}$$

$\rightsquigarrow$ Slowly decaying Kolomogorov N-width (Greif,Urban,2019)

(Discrete) Hamiltonian form: canonical, quadratic

$$\mathcal{H}(\mathbf{x};\boldsymbol{\mu}) = \frac{1}{2}\mathbf{x}^T \begin{bmatrix} -\left(\frac{\boldsymbol{\mu}}{\Delta\xi}\right)^2 D & 0 \\ 0 & I \end{bmatrix} \mathbf{x}$$

- Time discretization: implicit midpoint with $n_t = 4000$
- Training data: $|\mathcal{P}_{\text{train}}| = 8 \rightsquigarrow 32000$ snapshots

Weakly Symplectic Deep Convolutional Autoencoder

Loss function:

$$\mathcal{L}(\boldsymbol{\theta}) := \alpha \mathcal{L}_{\mathsf{data}}(\boldsymbol{\theta}) + (1-\alpha)\mathcal{L}_{\mathsf{sympl}}(\boldsymbol{\theta}),$$

with data and symplecticity loss

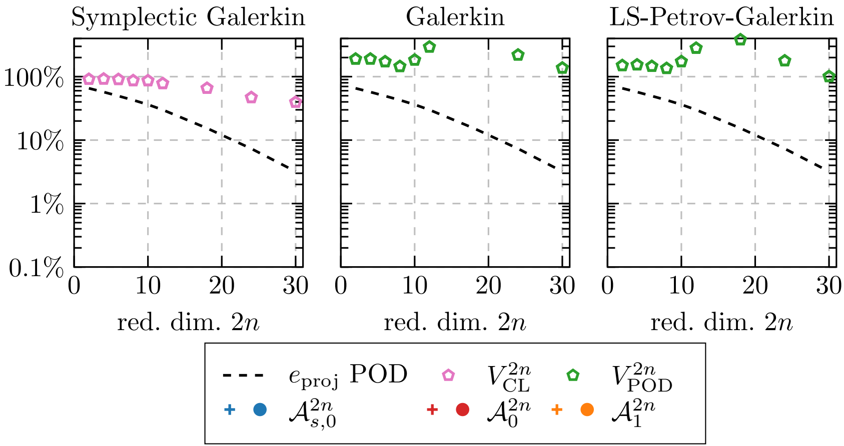
$$\mathcal{L}_{\mathsf{data}}(\boldsymbol{\theta}) := \frac{1}{N|\mathcal{X}|} \sum_{\mathbf{x}\in\mathcal{X}} \|\mathbf{x} - \varphi(\varrho(\mathbf{x};\boldsymbol{\theta});\boldsymbol{\theta})\|^2,$$

$$\mathcal{L}_{\mathsf{sympl}}(\boldsymbol{\theta}) := \frac{1}{(2n)^2|\mathcal{X}|} \left\| \left(\mathbf{D}_{\check{\mathbf{x}}}\varphi(\cdot;\boldsymbol{\theta})\big|_{\varrho(\mathbf{x};\boldsymbol{\theta})} \right)^{\mathsf{T}} \mathbb{J}_{2N}\, \mathbf{D}_{\check{\mathbf{x}}}\varphi(\cdot;\boldsymbol{\theta})\big|_{\varrho(\mathbf{x};\boldsymbol{\theta})} - \mathbb{J}_{2n} \right\|_{\mathsf{F}}^2,$$

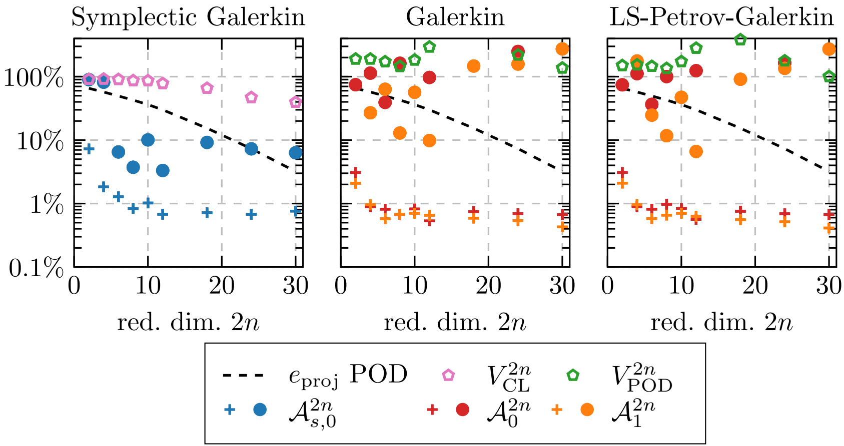
Autocoder architectures: basis sizes $2n \in \{2, 4, \ldots, 12, 18, 24, 30\}$

- weakly symplectic DCA $\mathcal{A}_{s,0}^{2n}$ architecture,
- a non-symplectic copy $\mathcal{A}_0^{2n}$,
- completely different non-symplectic DCA architecture $\mathcal{A}_1^{2n}$.

Numerical Results: Projection Error/Reduction Error

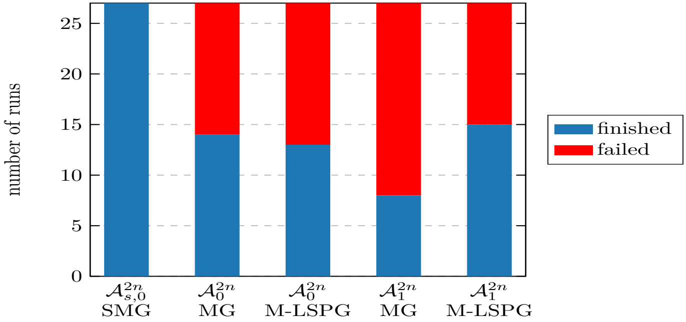


Numerical Results: Projection Error/Reduction Error



Numerical Results: Reduced Solution Runs

- Number of successful and failed reduced simulation runs,
- for $\mathcal{A}^{2n}_{s,0}, \mathcal{A}^{2n}_{0}, \mathcal{A}^{2n}_{1}$ with respective reduction technique,
- for μ_1, μ_2, μ_3 and $2n \in \{2, 4, \ldots, 12, 18, 24, 30\}$.



Outlook:

1. Introduction to classical Model Reduction
2. Model Reduction on Manifolds
3. Computation of Reduced Model from Data
4. Conclusion

Conclusion & Extensions

Conclusion:[1]

- Presented framework of model reduction on manifolds
- Structure-preservation with GMG
- Reduction of Hamiltonian system

Extensions:

- Extending of formulation to port-Hamiltonian systems
- Including symplecticity directly in the architecture of the autoencoder

[1] P. Buchfink, SG., B. Haasdonk, B. Unger. *Model Reduction on Manifolds: a differential geometric framework*, arXiv:2312.01963, 2023.