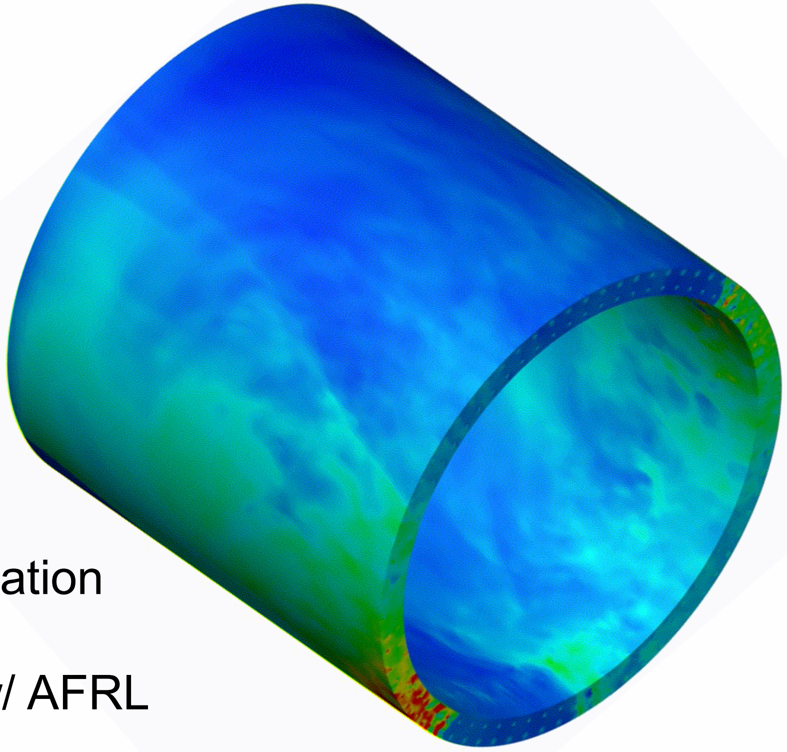
Learning structure-exploiting reduced models with Operator Inference

Professor Karen E. Willcox
Green Family Lecture
Institute for Pure and Applied Mathematics
April 14, 2026

ODEN INSTITUTE
FOR COMPUTATIONAL ENGINEERING & SCIENCES
TEXAS
The University of Texas at Austin

Physics-based models are powerful and bring predictive capabilities

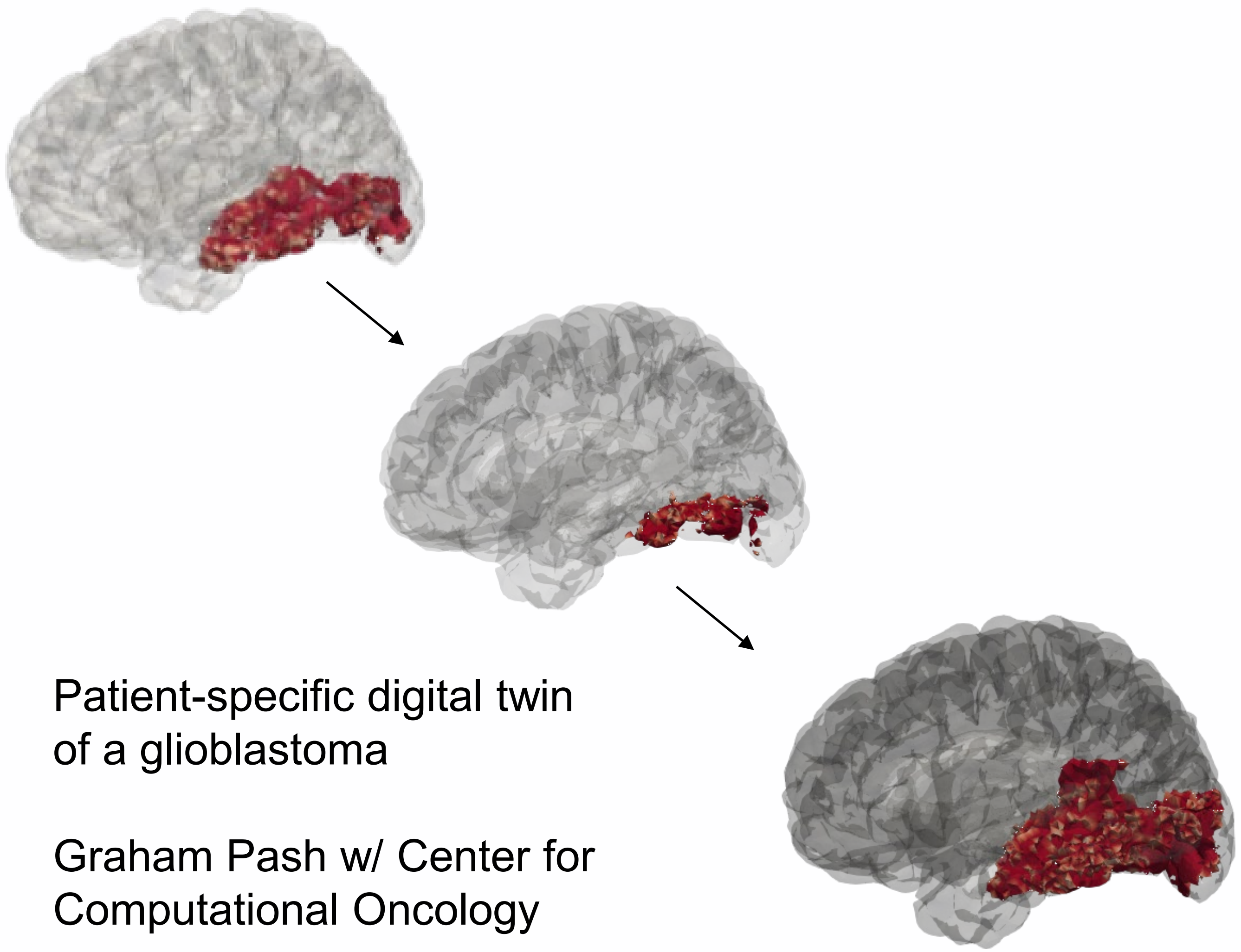


Rotating detonation rocket engine
Ionut Farcas w/ AFRL

Large eddy simulation (LES) of reactive Navier-Stokes equations with 136M spatial dof and $\Delta t = 10^{-9}$

but they can be **COMPUTATIONALLY PROHIBITIVE** for design, control, and UQ

Physics-based models are powerful and bring predictive capabilities

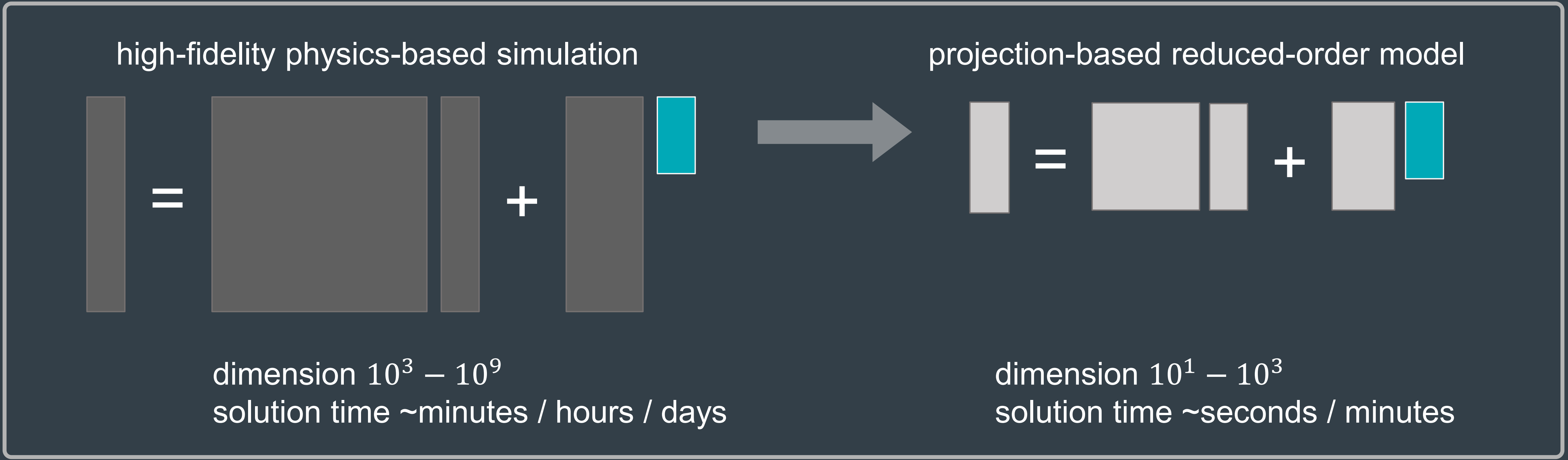


Patient-specific digital twin of a glioblastoma

Graham Pash w/ Center for Computational Oncology

but they can be COMPUTATIONALLY PROHIBITIVE for digital twins

Reduced-order models are critical enablers for design, control, UQ and digital twins



1 **Train**: Solve PDEs to generate training data (snapshots)

2 **Identify structure**: Identify a low-dimensional manifold

3 **Reduce**: Project PDE model onto the low-dimensional manifold

Why not use a neural operator?

- Projection-based model reduction preserves physics structure by construction

- Projection-based reduced models are interpretable (evolution of modal coordinates)

- Projection-based reduced models can be learned directly from data (non-intrusive)

Model reduction meets machine learning

Machine learning

“Machine learning is a field of study in artificial intelligence concerned with the development and study of statistical algorithms that can learn from data and generalize to unseen data, and thus perform tasks without explicit programming language instructions.” [Wikipedia]

Reduced-order modeling

“Model order reduction is a technique for reducing the computational complexity of mathematical models in numerical simulations. As such it is closely related to the concept of metamodeling, with applications in all areas of mathematical modelling.” [Wikipedia]

ML surrogates vs. model reduction

Machine learning

"…statistical algorithms that can learn from data and generalize to unseen data, and thus perform tasks without explicit programming language instructions." [Wikipedia]

Reduced-order modeling

"Model order reduction is a technique for reducing the computational complexity of mathematical models in numerical simulations." [Wikipedia]

Model reduction methods have grown from Computational Science, with focus on *reducing* high-dimensional models that arise from physics-based modeling, whereas machine learning has grown from Computer Science, with focus on *learning* models from black-box data streams.
[Swischuk et al., *Computers & Fluids*, 2019]

ML surrogates vs. model reduction

Machine learning

- Models may or may not generalize
- Large training data requirements
- Non-intrusive, portable & flexible
- Accessible & available
- Massive uptake outside the expert community

Reduced-order modeling

- Deep body of theory & methods, (stability, structure preservation, error estimators, …)
- Highly expert community
- Methods are inaccessible & intrusive
- Limited uptake outside the expert community

Available online at www.sciencedirect.com
ScienceDirect

Comput. Methods Appl. Mech. Engrg. 306 (2016) 196–215

Computer methods in applied mechanics and engineering

www.elsevier.com/locate/cma

Data-driven operator inference for nonintrusive projection-based model reduction

Benjamin Peherstorfer*, Karen Willcox

Department of Aeronautics & Astronautics, MIT, 77 Massachusetts Avenue, Cambridge, MA 02139, USA

Received 7 September 2015; received in revised form 2 February 2016
Available online 6 April 2016

Abstract

This work presents a nonintrusive projection-based model reduction approach for full models based on time-dependent partial differential equations. Projection-based model reduction constructs the operators of a reduced model by projecting the equations of the full model onto a reduced space. Traditionally, this projection is intrusive, which means that the full-model operators are required either explicitly in an assembled form or implicitly through a routine that returns the action of the operators on a given vector; however, in many situations the full model is given as a black box that computes trajectories of the full-model states and outputs for given initial conditions and inputs, but does not provide the full-model operators. Our nonintrusive operator inference approach infers approximations of the reduced operators from the initial conditions, inputs, trajectories of the states, and outputs of the full model, without requiring the full-model operators. Our operator inference is applicable to full models that are linear in the state or have a low-order polynomial nonlinear term. The inferred operators are the solution of a least-squares problem and converge, with sufficient state trajectory data, in the Frobenius norm to the reduced operators that would be obtained via an intrusive projection of the full-model operators. Our numerical results demonstrate operator inference on a linear climate model and on a tubular reactor model with a polynomial nonlinear term of third order.

Keywords: Nonintrusive model reduction; Data-driven model reduction; Black-box full model; Inference

1. Introduction

Model reduction seeks to construct reduced models that provide accurate approximations of the full model solutions with orders of magnitude reduction in computational complexity. We consider here projection-based model reduction

The Operator Inference problem

Given (1) a physical/natural system with known governing equations, and (2) a set of data in the form of state snapshots (experimental or simulation)

Infer a reduced-order model that recovers the given data and provides a predictive capability to rapidly simulate unseen conditions

$$\min_{\mathbf{O}} ||\mathbf{D}\mathbf{O} - \mathbf{R}||$$

$\mathbf{O}$: low-dimensional operators that define the reduced model

$\mathbf{D}, \mathbf{R}$: data matrix / forcing from simulation and/or experimental data

We use:

- the **physics** to define the structured form of the model we seek
- **projection-based model reduction** to cast the inference in a reduced coordinate space and to provide error estimates in some settings
- **inverse theory** to analyze the structure of the resulting problem and treat it numerically
- **numerical linear algebra** to achieve efficient scalable algorithms

The Operator Inference ROM form is defined by projection-based reduction theory

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u}$$

Full-order model (FOM)
state $\mathbf{x} \in \mathbb{R}^N$
input $\mathbf{u} \in \mathbb{R}^{N_i}$

Approximate
$\mathbf{x} \approx \mathbf{V}\hat{\mathbf{x}}$
$\mathbf{V} \in \mathbb{R}^{N \times r}$

Inside-Out View

Residual: N eqs $\gg r$ dof

$$\mathbf{r} = \mathbf{V}\dot{\hat{\mathbf{x}}} - \mathbf{A}\mathbf{V}\hat{\mathbf{x}} - \mathbf{B}\mathbf{u}$$

Project
$\mathbf{W}^\top \mathbf{r} = 0$
(Galerkin: $\mathbf{W} = \mathbf{V}$)

In classical projection approaches:

$$\widehat{\mathbf{A}} = \mathbf{V}^\top \mathbf{A}\mathbf{V}$$
$$\widehat{\mathbf{B}} = \mathbf{V}^\top \mathbf{B}$$

$$\dot{\hat{\mathbf{x}}} = \widehat{\mathbf{A}}\hat{\mathbf{x}} + \widehat{\mathbf{B}}\mathbf{u}$$

Reduced-order model (ROM)
state $\hat{\mathbf{x}} \in \mathbb{R}^r,\ r \ll N$

OPERATOR INFERENCE reflects the structure of a physics-based model to learn reduced models from data

full-order model (FOM):

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u}$$

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{H}(\mathbf{x} \otimes \mathbf{x}) + \mathbf{B}\mathbf{u}$$

represent high-dimensional state $\mathbf{x} \in \mathbb{R}^N$ in a low dimensional basis $\mathbf{V} \in \mathbb{R}^{N \times r}$: $\mathbf{x} \approx \mathbf{V}\mathbf{x}_r$

reduced-order model (ROM):

$$\dot{\hat{\mathbf{x}}} = \hat{\mathbf{A}}\hat{\mathbf{x}} + \hat{\mathbf{B}}\mathbf{u}$$

$$\dot{\hat{\mathbf{x}}} = \hat{\mathbf{A}}\hat{\mathbf{x}} + \hat{\mathbf{H}}(\hat{\mathbf{x}} \otimes \hat{\mathbf{x}}) + \hat{\mathbf{B}}\mathbf{u}$$

The ROM form is inspired by classical intrusive physics-based model reduction, but the operators are learned directly from data

Our Operator Inference approach blends model reduction & machine learning

Define the **structure of the reduced model**

$$\dot{\hat{\mathbf{x}}} = \hat{\mathbf{A}}\hat{\mathbf{x}} + \hat{\mathbf{B}}\mathbf{u} + \hat{\mathbf{H}}(\hat{\mathbf{x}} \otimes \hat{\mathbf{x}})$$

Non-intrusive learning by inferring reduced operators from simulation data [Peherstorfer & W., 2016]

snapshots generated from projected simulation data

low-dimensional operators define the reduced model as a dynamical system

$$\min_{\hat{\mathbf{A}},\hat{\mathbf{B}},\hat{\mathbf{H}}} \left\| \hat{\mathbf{X}}^\top \hat{\mathbf{A}}^\top + \left(\hat{\mathbf{X}} \otimes \hat{\mathbf{X}}\right)^\top \hat{\mathbf{H}}^\top + \mathbf{U}^\top \hat{\mathbf{B}}^\top - \dot{\hat{\mathbf{X}}}^\top \right\|$$

minimum residual formulation leads to linear least squares

- For Markovian data, OpInf has preasymptotic **recovery of intrusive ROM** [Peherstorfer, 2020]
- Projection-based model reduction **preserves physics structure** by construction
- Projection-based reduced models are **interpretable** (evolution of modal coordinates)

Our Operator Inference approach blends model reduction & machine learning

Define the **structure of the reduced model**

$$\dot{\hat{\mathbf{x}}} = \widehat{\mathbf{A}}\hat{\mathbf{x}} + \widehat{\mathbf{B}}\mathbf{u} + \widehat{\mathbf{H}}(\hat{\mathbf{x}} \otimes \hat{\mathbf{x}})$$

Non-intrusive learning by inferring reduced operators from simulation data [Peherstorfer & W., 2016]

Operator Inference is non-intrusive and scalable

1. Generate snapshots from high-fidelity simulation
2. Compute POD basis (SVD) – we have the freedom to choose our variables
3. Compute snapshot low-dimensional representation (matrix-vector multiplications)
4. Solve linear least squares minimization problem to infer the low-dimensional model

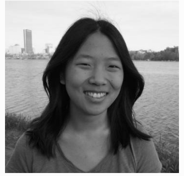
Our Operator Inference approach blends model reduction & machine learning

Define the **structure of the reduced model**

$$\dot{\hat{\mathbf{x}}} = \hat{\mathbf{A}}\hat{\mathbf{x}} + \hat{\mathbf{B}}\mathbf{u} + \hat{\mathbf{H}}(\hat{\mathbf{x}} \otimes \hat{\mathbf{x}})$$

Non-intrusive learning by inferring reduced operators from simulation data [Peherstorfer & W., 2016]

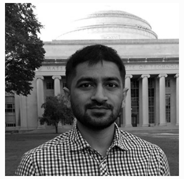
- OpInf for non-polynomial nonlinear systems [Qian et al., 2020, Benner et al., 2020]
- Regularization is key [McQuarrie, Huang & W., 2021]
- OpInf formulation in PDE setting [Qian, Farcas & W., 2022]
- Parametric OpInf [McQuarrie, S., Khodabakhshi & W., 2023]
- Quadratic Manifold OpInf [Geelen, Wright & W., 2023, Geelen et al. 2024]
- Bayesian OpInf [Guo, McQuarrie & W., 2022]
- Distributed OpInf [Farcas et al., 2025]
- Block-structured OpInf [Zastrow et al., 2025]
- Nested OpInf [Aretz & W., 2025]










Distributed Operator Inference

enables reduced modeling at scale

Given snapshots from high-fidelity simulation (or experiment)
generate projected transformed snapshots

1. Load snapshots from disk

$$\mathbf{X}_{\mathbf{orig}} = \begin{bmatrix} | & & | \\ \mathbf{x}(t_1) & \dots & \mathbf{x}(t_K) \\ | & & | \end{bmatrix}$$

2. Apply data transformations, such as lifting, centering & scaling

$$\mathbf{X}_{\mathbf{orig}} \longrightarrow \mathbf{X}$$

3. Compute POD basis

$$\mathbf{X} = \mathbf{V}\,\mathbf{\Sigma}\,\mathbf{W}^\top$$

4. Compute global representation of transformed data in low-dimensional space

$$\hat{\mathbf{X}} = \mathbf{V}^\top \mathbf{X}$$

Given low-dimensional snapshot data,
infer reduced model operators

$$\min_{\hat{\mathbf{A}},\hat{\mathbf{B}},\hat{\mathbf{H}}} \left\| \hat{\mathbf{X}}^\top \hat{\mathbf{A}}^\top + \left(\hat{\mathbf{X}} \otimes \hat{\mathbf{X}}\right)^\top \hat{\mathbf{H}}^\top + \mathbf{U}^\top \hat{\mathbf{B}}^\top - \dot{\hat{\mathbf{X}}}^\top \right\|_F^2 + \text{regularization}$$

1. Loop over regularization parameters; solve regularized OpInf least-squares problem

Distributed Operator Inference
enables reduced modeling at scale

Farcas, Gundevia, Munipalli & W., 2024

Given snapshots from high-fidelity simulation (or experiment)
generate projected transformed snapshots

1. Load snapshots from disk
2. Apply data transformations, such as lifting, centering & scaling
3. Compute POD basis
4. Compute global representation of transformed data in low-dimensional space

partition snapshots across p processing units; choose partitions to minimize communication in computing data transformations

use POD method of snapshots:

$$\mathbf{C} = \mathbf{X}^\top \mathbf{X} \qquad \mathbf{T}_r = \mathbf{U}_r \boldsymbol{\Lambda}_r^{-1/2}$$

$$\mathbf{C}\mathbf{U} = \mathbf{U}\boldsymbol{\Lambda} \qquad \hat{\mathbf{X}} = \mathbf{T}_r \mathbf{C}$$

Given low-dimensional snapshot data,
infer reduced model operators

1. Loop over regularization parameters; solve regularized OpInf least-squares problem

split regularization parameters into disjoint subsets; perform parallel reduction to find optimal regularization parameters

Different variable choices lead to different **structure** in the system

Example: There are multiple ways to write the Euler equations

$$\frac{\partial}{\partial t}\begin{pmatrix}\rho\\ \rho w\\ E\end{pmatrix}+\frac{\partial}{\partial z}\begin{pmatrix}\rho w\\ \rho w^2+p\\ (E+p)w\end{pmatrix}=0$$

$$E=\frac{p}{\gamma-1}+\frac{1}{2}\rho w^2$$

conservative variables
mass, momentum, energy

$$\frac{\partial}{\partial t}\begin{pmatrix}\rho\\ w\\ p\end{pmatrix}+\begin{pmatrix}\rho\frac{\partial w}{\partial z}+w\frac{\partial \rho}{\partial z}\\ w\frac{\partial w}{\partial z}+\frac{1}{\rho}\frac{\partial p}{\partial z}\\ \gamma p\frac{\partial w}{\partial z}+w\frac{\partial p}{\partial z}\end{pmatrix}=0$$

primitive variables
mass, velocity, pressure

$$\frac{\partial}{\partial t}\begin{pmatrix}w\\ p\\ q\end{pmatrix}+\begin{pmatrix}w\frac{\partial w}{\partial z}+q\frac{\partial p}{\partial z}\\ \gamma p\frac{\partial w}{\partial z}+w\frac{\partial p}{\partial z}\\ q\frac{\partial w}{\partial z}+w\frac{\partial q}{\partial z}\end{pmatrix}=0$$

specific volume variables

$$q={}^{1}/_{\rho}$$

↔

$$\dot{\mathbf{x}}=\mathbf{H}(\mathbf{x}\otimes\mathbf{x})+\mathbf{B}\mathbf{u}$$

transformed system
has quadratic structure

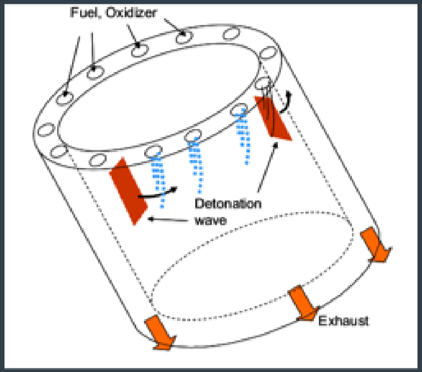
↕

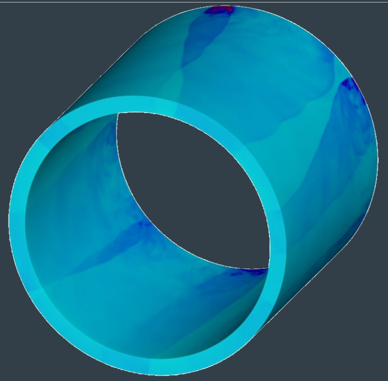
$$\dot{\mathbf{x}}_r=\mathbf{H}_r(\mathbf{x}_r\otimes\mathbf{x}_r)+\mathbf{B}_r\mathbf{u}$$

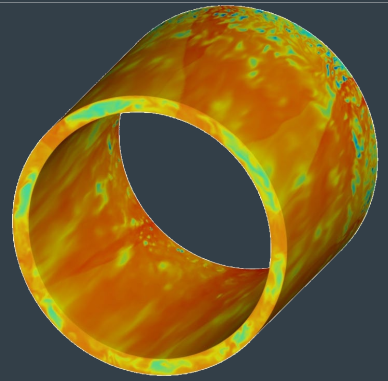
ROM has quadratic structure

Modeling the combustion chamber of a rotating detonation rocket engine

- LES simulations of the reactive, viscous 3D Navier-Stokes equations
- Skeletal chemistry mechanism based on the Foundational Fuel Chemistry Model (FFCMy-12)
- Non-premixed fuel injection (gaseous methane and oxygen)





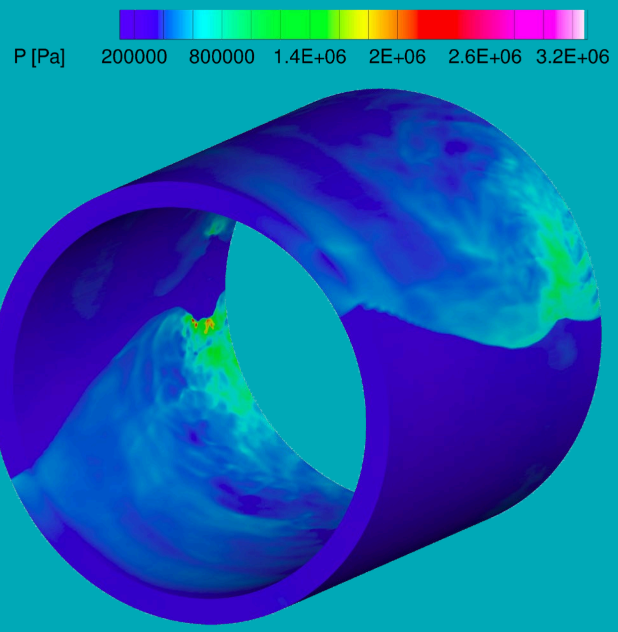


$$\frac{\partial}{\partial t}\begin{bmatrix}\rho\\ \rho v_x\\ \rho v_y\\ \rho E\\ \rho Y_1\\ \vdots\\ \rho Y_{n_{sp}}\end{bmatrix}+\nabla\cdot\left(\begin{bmatrix}\rho v_x\\ \rho v_x^2+p\\ \rho v_x v_y\\ \rho v_x E+p v_x\\ \rho v_x Y_1\\ \vdots\\ \rho v_x Y_{n_{sp}}\end{bmatrix}\vec{i}+\begin{bmatrix}\rho v_y\\ \rho v_x v_y\\ \rho v_y^2+p\\ \rho v_y E+p v_y\\ \rho v_y Y_1\\ \vdots\\ \rho v_y Y_{n_{sp}}\end{bmatrix}\vec{j}-\begin{bmatrix}0\\ \tau_{xx}\\ \tau_{yx}\\ \tau_{xx}v_x+\tau_{yx}v_y-j_x^q\\ -j_{1,x}^m\\ \vdots\\ -j_{n_{sp},x}^m\end{bmatrix}\vec{i}-\begin{bmatrix}0\\ \tau_{xy}\\ \tau_{yy}\\ \tau_{xy}v_x+\tau_{yy}v_y-j_y^q\\ -j_{1,y}^m\\ \vdots\\ -j_{n_{sp},y}^m\end{bmatrix}\vec{j}\right)=\begin{bmatrix}0\\ 0\\ 0\\ 0\\ \dot{\omega}_1\\ \vdots\\ \dot{\omega}_{n_{sp}}\end{bmatrix}$$

Modeling the combustion chamber of a rotating detonation rocket engine

1. Training data & data transformations

- LES simulation 136M spatial dof for full RDRE, timestep $\Delta t = 10^{-9}$s, ~1M CPU hours per 1ms on 16K cores
- Available data: 501 down-sampled combustion chamber snapshots over [2.50, 3.75] ms (~4 periods of two-wave system) interpolated onto structured grid with 4.2M dof
- 18 transformed state variables (+ scale & center): specific volume, pressure, 3D velocity, temperature, 12 species mass fractions (full chemistry data)
- Transformed training data: snapshots $\mathbf{X} \in \mathbb{R}^{76M\times 375}$



Two dominant co-rotating waves in the quasi-limit-cycle behavior of the flow

Modeling the combustion chamber of a rotating detonation rocket engine

2. Compute low-dimensional representation

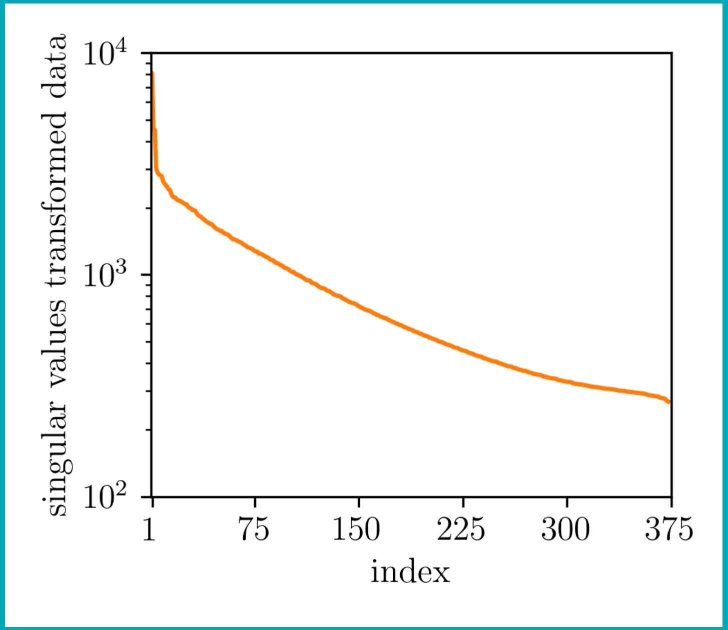
- Snapshot matrix of transformed variables

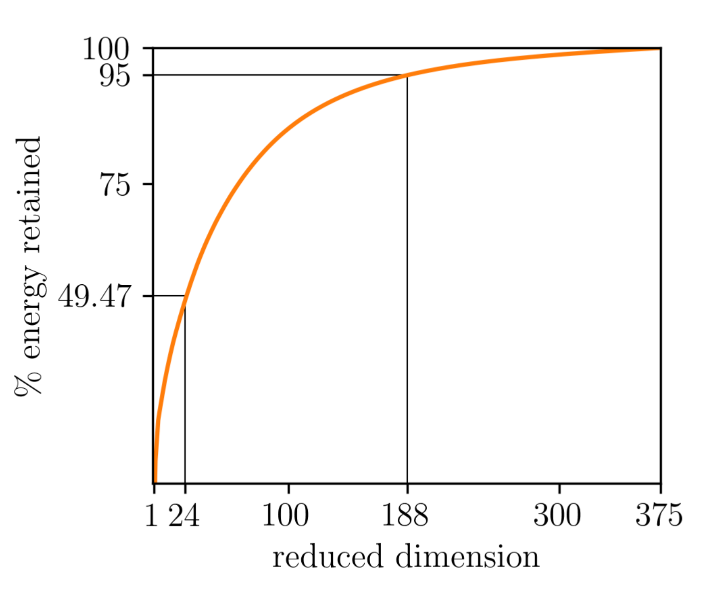
$$\mathbf{X} \in \mathbb{R}^{76\text{M}\times 375}$$

$$\widehat{\mathbf{X}} = \mathbf{U}_r \mathbf{\Lambda}_r^{-1/2} \mathbf{C} \in \mathbb{R}^{r\times 375}$$

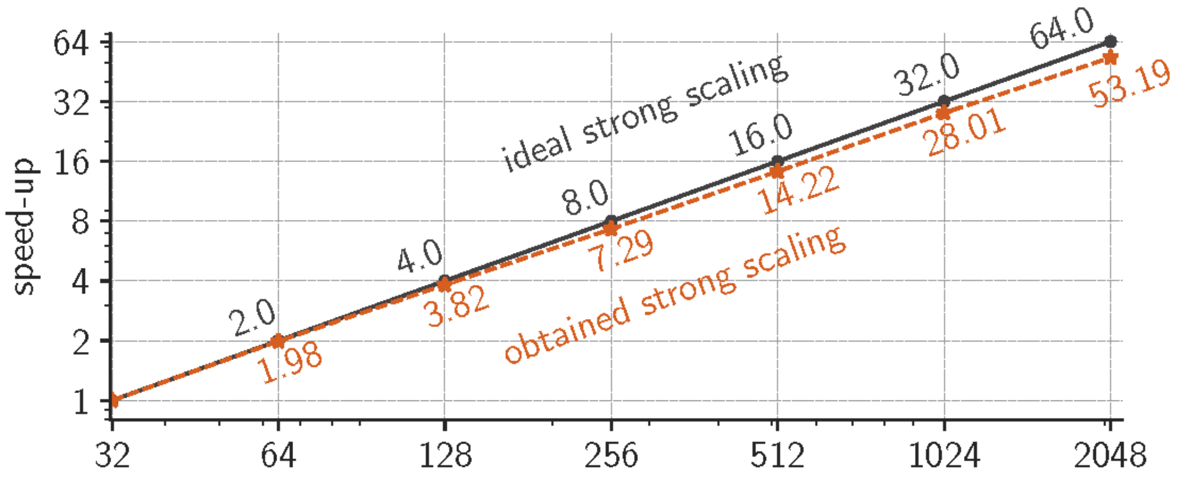
- Singular values guide the choice of r (low-data regime limits size of non-intrusive ROM)
- POD basis only computed if needed for reconstruction

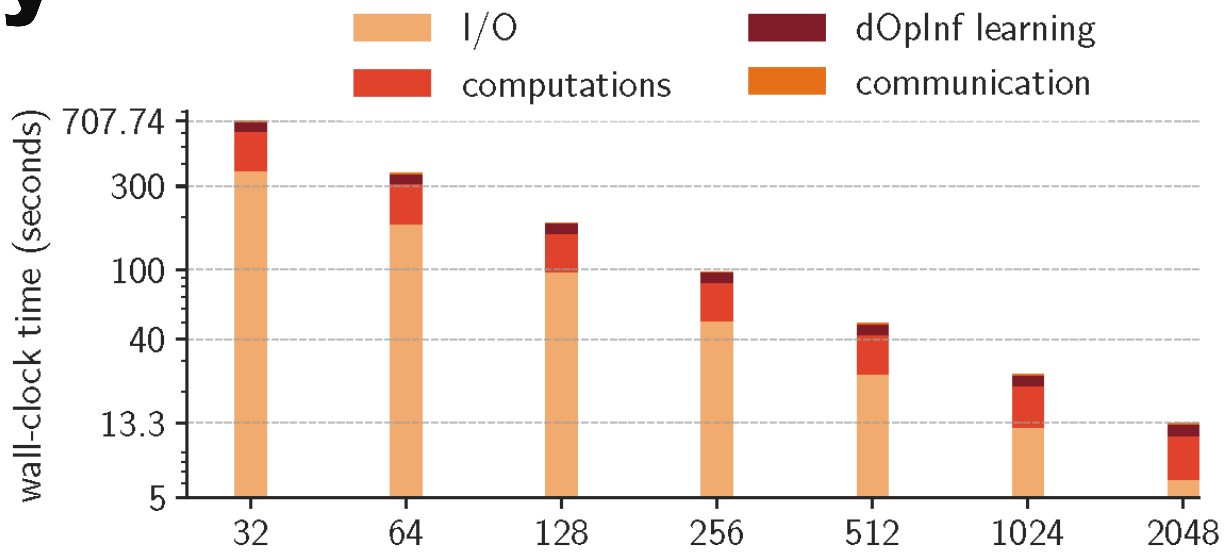
3. Infer reduced operators

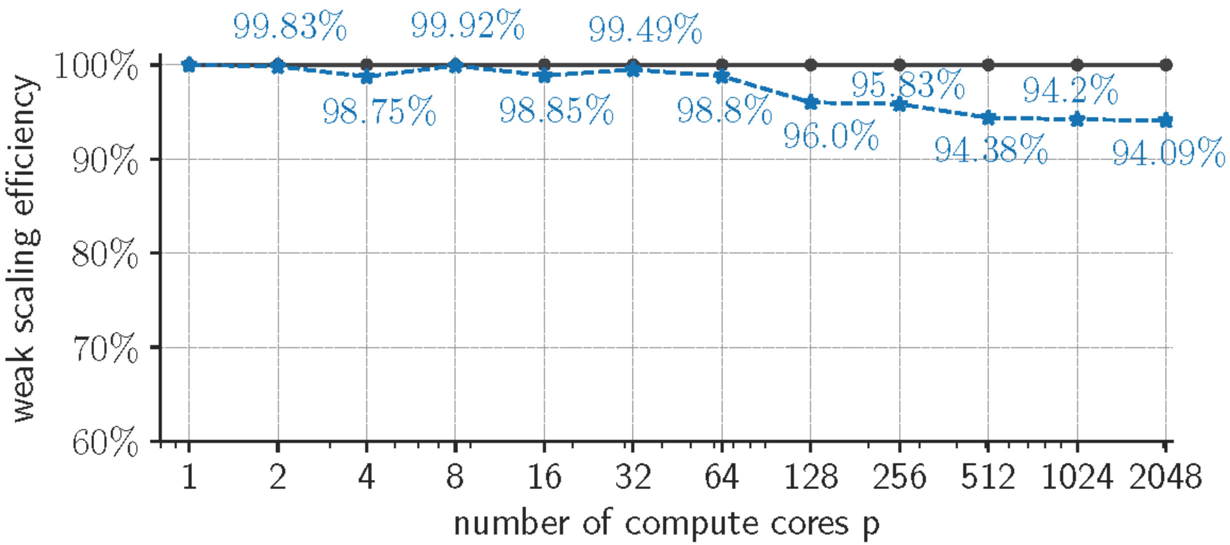


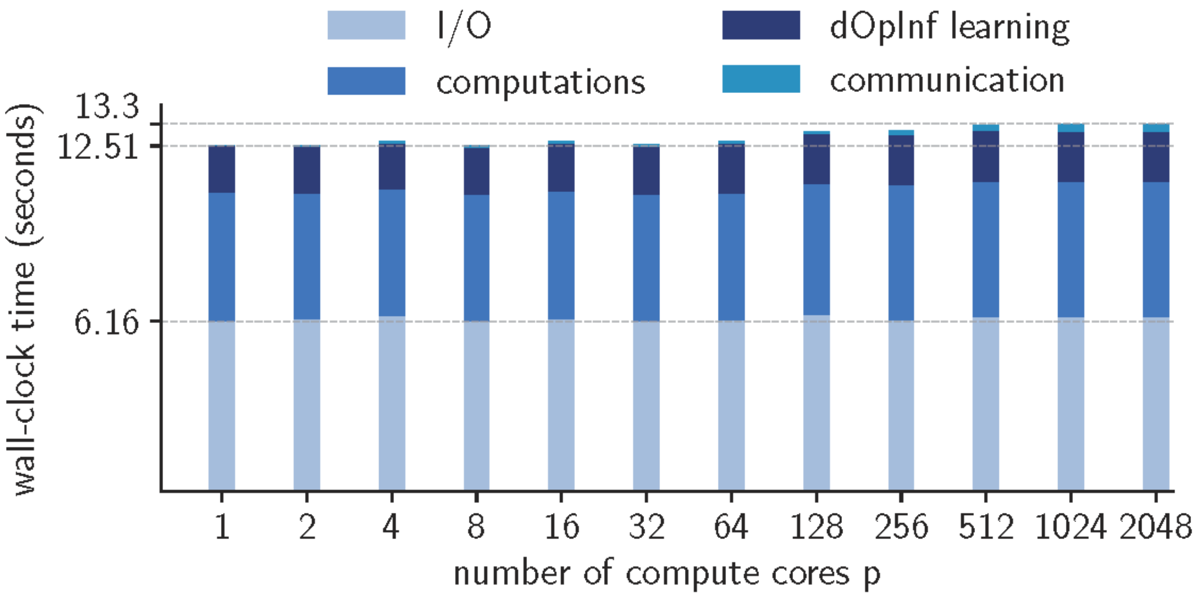


Parallel computations on Frontera (TACC) show excellent scalability



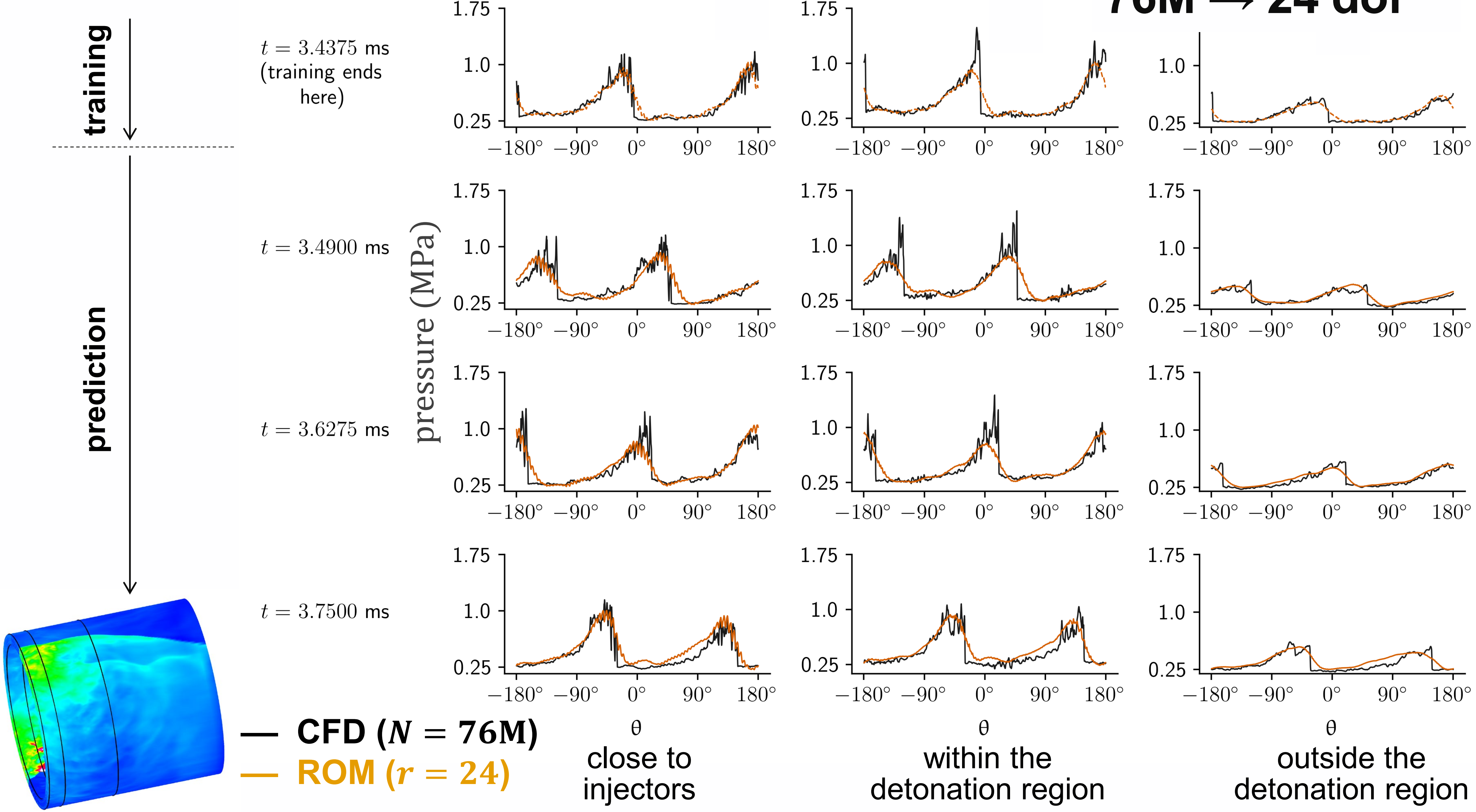


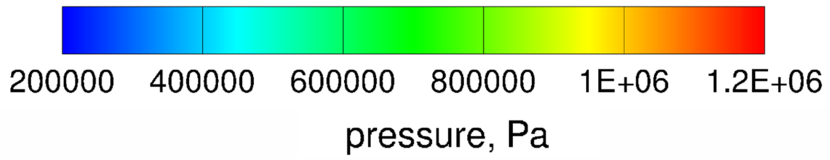




Rotating detonation rocket engine simulation: weeks → milliseconds

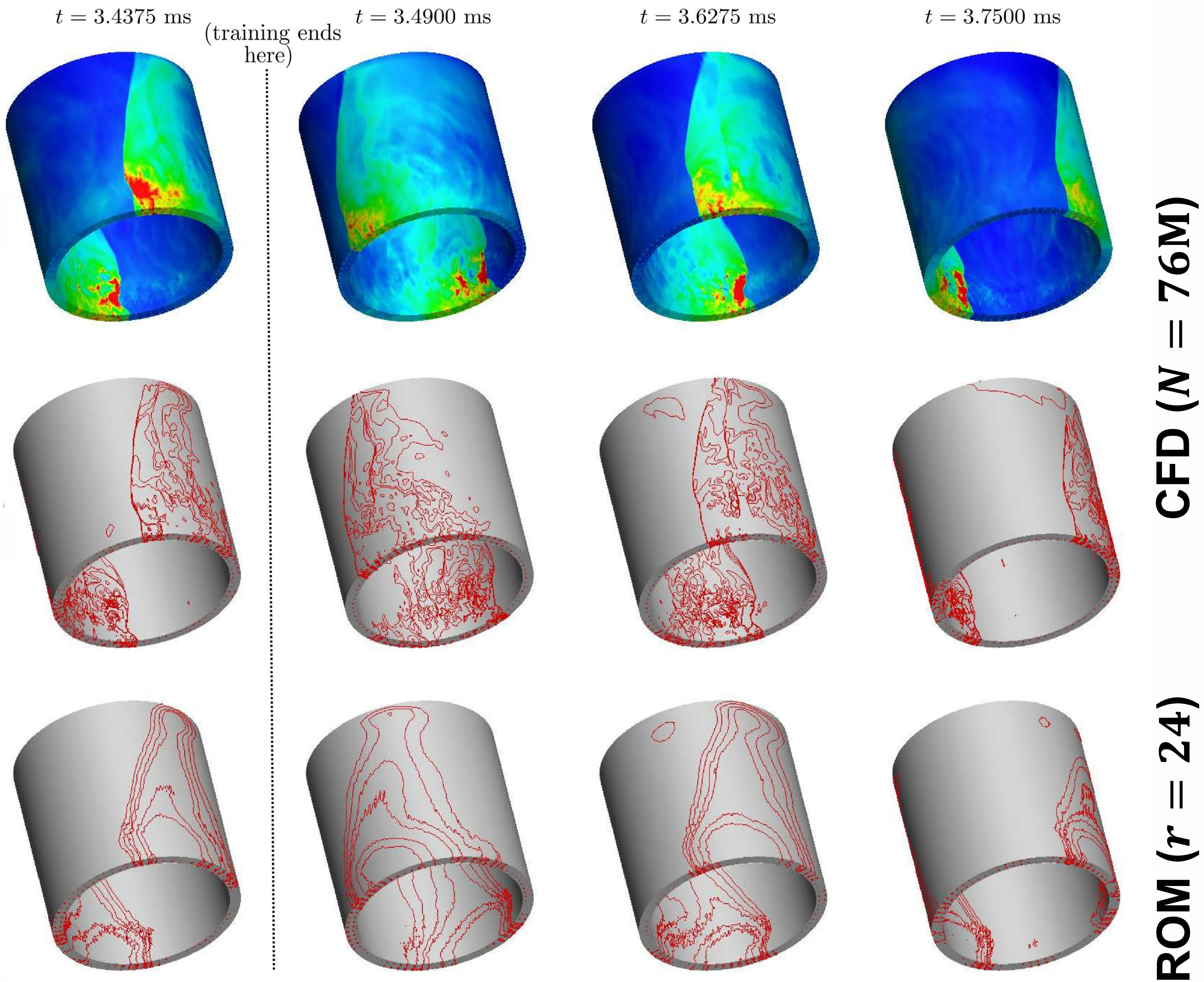
76M → 24 dof





RDRE pressure contours:

Reduced model captures coarse behavior but does not resolve all fine-scale dynamics

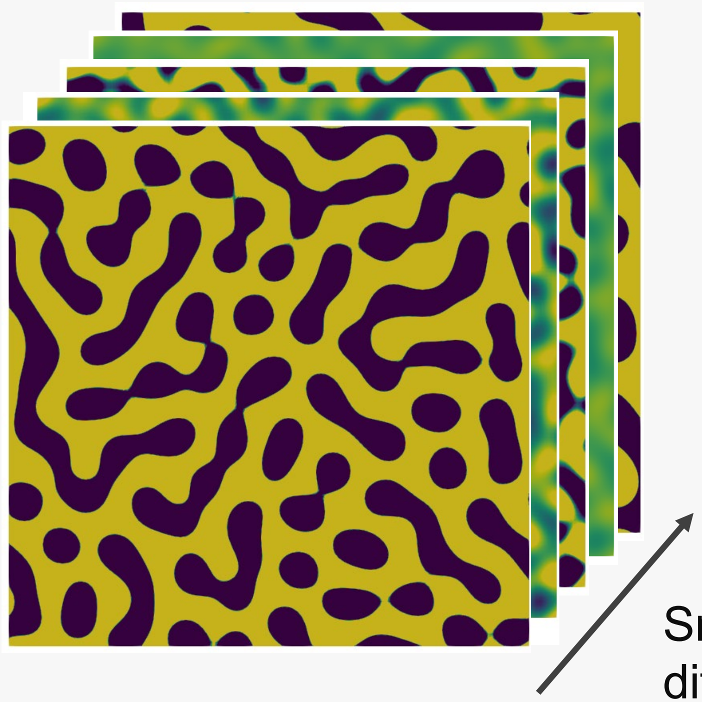


Sometimes a LINEAR BASIS **is NOT ENOUGH**

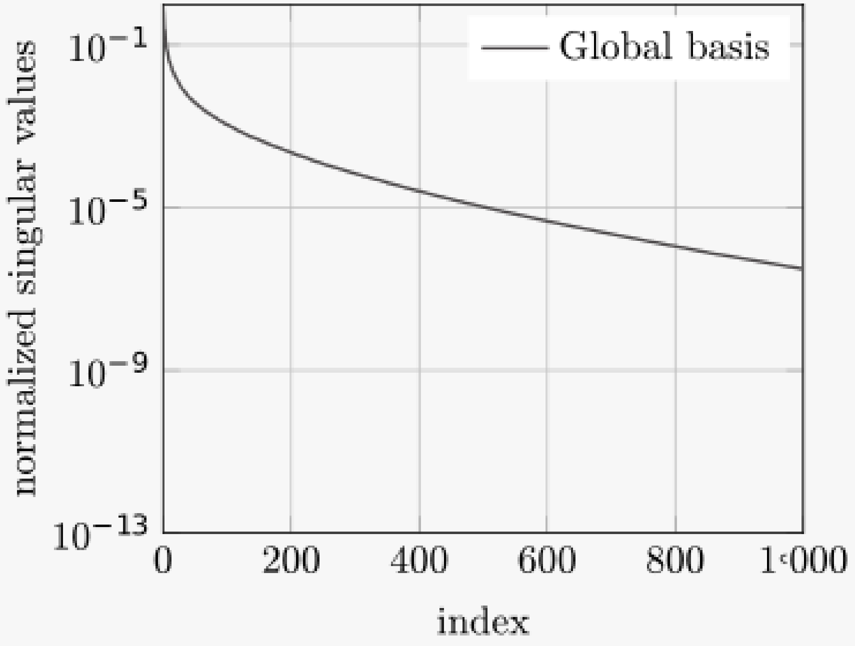
Rudy Geelen

Cahn-Hilliard phase-field model

$$\frac{\partial}{\partial t} s(\mathbf{x}, t) = M \nabla^2 \left(s^3(\mathbf{x}, t) - s(\mathbf{x}, t) - \ell \nabla^2 s(\mathbf{x}, t) \right)$$



Snapshot collection across different time steps and initial conditions

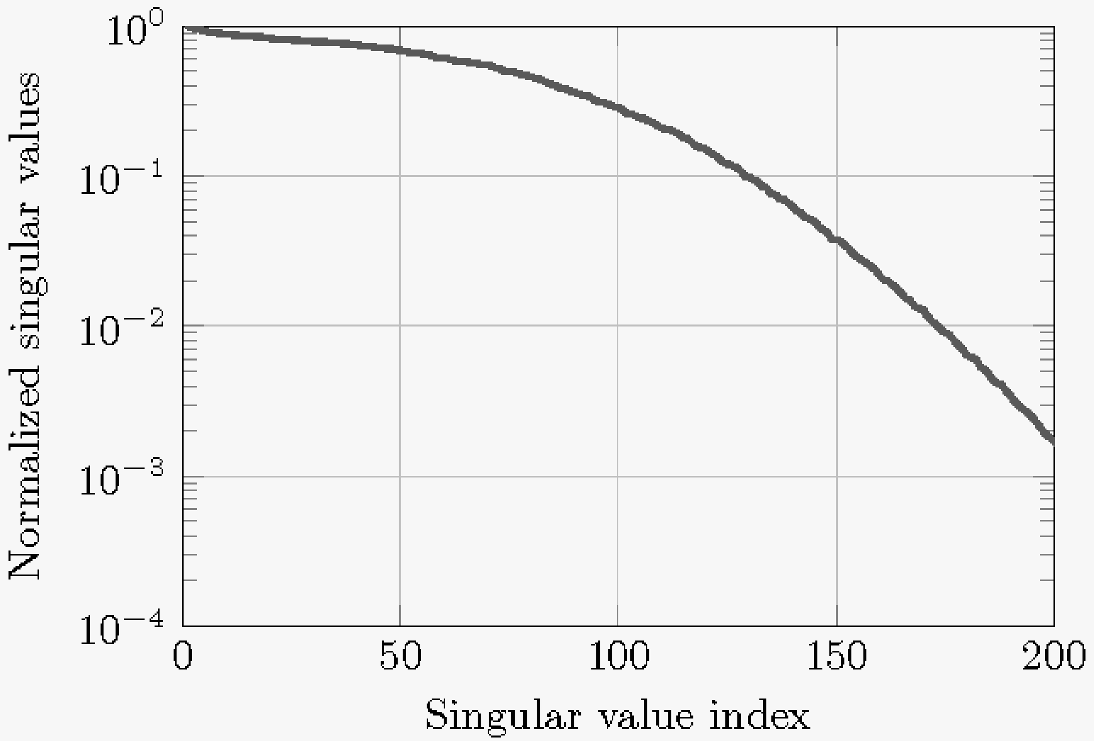


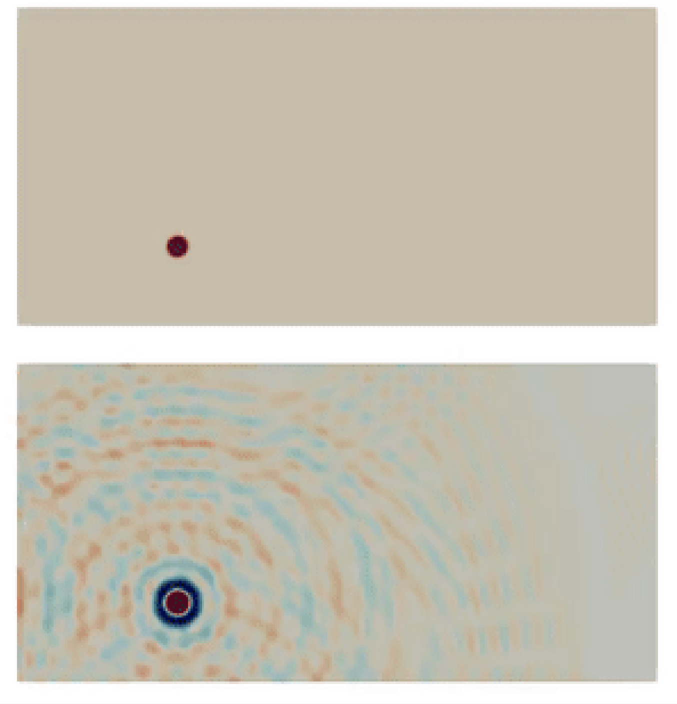
Slow decay of the singular values
→ reduced model has high dimension

Sometimes a LINEAR BASIS **is NOT ENOUGH**

Wave Equation

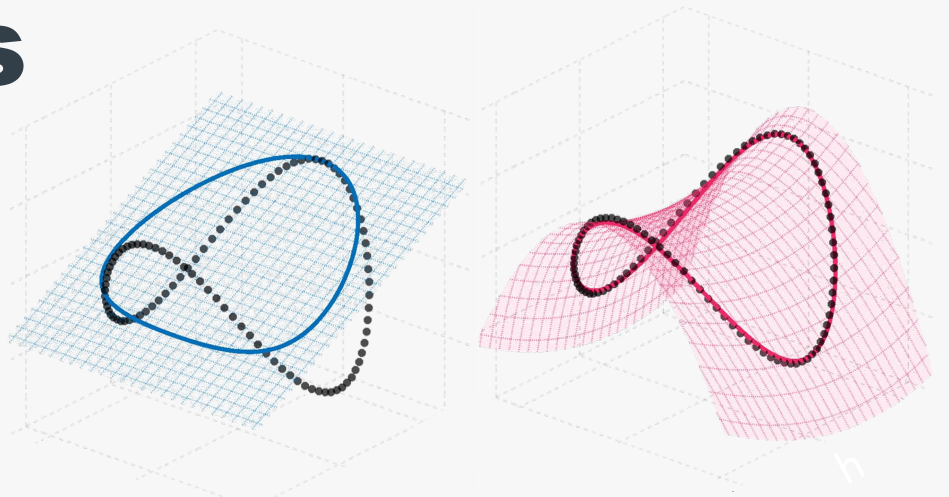
$$\frac{\partial^2}{\partial t^2} s(\mathbf{x}, t) = c^2 \Delta s$$





Go beyond a linear basis

via approximation in a nonlinear manifold



but maintain the structure-preserving properties of projection

Lift to quadratic form + linear projection

- Gu. *IEEE*, 2011
- Benner & Breiten. *SISC*, 2015
- Benner & Goyal. MTNS, 2016
- Benner, Goyal, Gugercin. *SIMAX*, 2018
- Kramer & W. *AIAA J.*, 2019

and many more

Approximation in a nonlinear manifold

- Rutzmoser, Rixen, Tiso, Jain. *Computers & Structures*, 2017
- Jain, Tiso, Rutzmoser, Rixen. *Computers & Structures*, 2017
- Barnett & Farhat. *JCP*, 2022
- Axås, Cenedese, Haller. *Nonlinear Dynamics*, 2022
- Barnett, Farhat, Maday. *JCP*, 2023
- Benner, Goyal, Heiland, Pontes Duff. *PAMM*, 2023
- Geelen, Wright, W. *CMAME*, 2023
- Geelen, Balzano, W. IEEE CDC, 2023
- Sharma, Mu, Buchfink, Geelen, Glas, Kramer. *CMAME*, 2023
- Cohen, Farhat, Maday, Somacal. C.R. Mecanique, 2023
- Geelen, Balzano, Wright, W. *Chaos*, 2024

Nonlinear dimensionality reduction via polynomial manifolds

polynomial terms are computed element-wise, i.e., if $\hat{\mathbf{x}} = (\hat{x}_1, \hat{x}_2)^\top$, then $\hat{\mathbf{x}}^p = \left(\hat{x}_1^p, \hat{x}_2^p\right)^\top$

Instead of the linear approximation $\mathbf{x} \approx \mathbf{V}\hat{\mathbf{x}}$, represent the high-dimensional state $\mathbf{x} \in \mathbb{R}^{\boldsymbol{N}}$ in a **polynomial manifold**:

$$\mathbf{x}(t) \approx \mathbf{x}_{\mathrm{ref}} + \mathbf{V}\hat{\mathbf{x}}(t) + \bar{\mathbf{V}}\boldsymbol{\Xi}\mathbf{g}\big(\hat{\mathbf{x}}(t)\big) \qquad \text{with} \quad \mathbf{g}(\hat{\mathbf{x}}) := \begin{pmatrix} \hat{\mathbf{x}}^2(t) \\ \hat{\mathbf{x}}^3(t) \\ \vdots \\ \hat{\mathbf{x}}^p(t) \end{pmatrix} \in \mathbb{R}^{(p-1)r}$$

- reference state $\mathbf{x}_{\mathrm{ref}} \in \mathbb{R}^N$ (scaling/shifting is important)
- reduced state $\hat{\mathbf{x}} \in \mathbb{R}^r$ (as before)
- basis matrices $\mathbf{V} \in \mathbb{R}^{N\times r}, \bar{\mathbf{V}} \in \mathbb{R}^{N\times q}$

- a coefficient matrix $\boldsymbol{\Xi} \in \mathbb{R}^{q\times(p-1)r}$
- polynomial terms up to degree $p \geq 2$ in vector $\mathbf{g}(\hat{\mathbf{x}}) \in \mathbb{R}^{(p-1)r}$ (other nonlinear terms possible too)

Nonlinear dimensionality reduction

How to compute $\mathbf{V}$, $\overline{\mathbf{V}}$, $\mathbf{\Xi}$, $\hat{\mathbf{x}}$?

- Given the polynomial manifold representation: $\mathbf{x}(t) \approx \mathbf{x}_{\text{ref}} + \mathbf{V}\hat{\mathbf{x}}(t) + \overline{\mathbf{V}}\mathbf{\Xi}\mathbf{g}\big(\hat{\mathbf{x}}(t)\big)$
- Choose the columns of $\mathbf{V}$ and $\overline{\mathbf{V}}$ to form an orthonormal set: $(\mathbf{V}, \overline{\mathbf{V}})^\top(\mathbf{V}, \overline{\mathbf{V}}) = \mathbf{I}_{\text{r+q}}$
- Given snapshot data $\mathbf{X} = \begin{pmatrix} | & | & & | \\ \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_k \\ | & | & & | \end{pmatrix}$,

  formulate a general **Representation Learning** problem:

$$\big(\mathbf{V}, \overline{\mathbf{V}}, \mathbf{\Xi}, \widehat{\mathbf{X}}\big) = \underset{\mathbf{V}, \overline{\mathbf{V}}, \mathbf{\Xi}, \widehat{\mathbf{X}}}{\operatorname{argmin}} \left(\frac{1}{2} \sum_{j=1}^{k} \left\| \mathbf{x}_j - \mathbf{x}_{\text{ref}} - (\mathbf{V} \quad \overline{\mathbf{V}}) \begin{pmatrix} \hat{\mathbf{x}}_j \\ \mathbf{\Xi}\mathbf{g}\big(\hat{\mathbf{x}}_j\big) \end{pmatrix} \right\|_{\mathbf{2}}^{\mathbf{2}} + \frac{\gamma}{2} \|\mathbf{\Xi}\|_{\text{F}}^2 \right)$$

$$\text{such that } (\mathbf{V}\,, \overline{\mathbf{V}})^\top(\mathbf{V}, \overline{\mathbf{V}}) = \mathbf{I}_{\text{r+q}}$$

Nonlinear Dimensionality Reduction
Simple illustrative example #1

3D trajectory parameterized by variable $t \in [0,2\pi]$

$$\mathbf{x}(t) = \begin{pmatrix} x_1(t) \\ x_2(t) \\ x_3(t) \end{pmatrix} = \begin{pmatrix} \cos(t) \\ \sin(t) \\ \cos(2t)/2 \end{pmatrix}$$

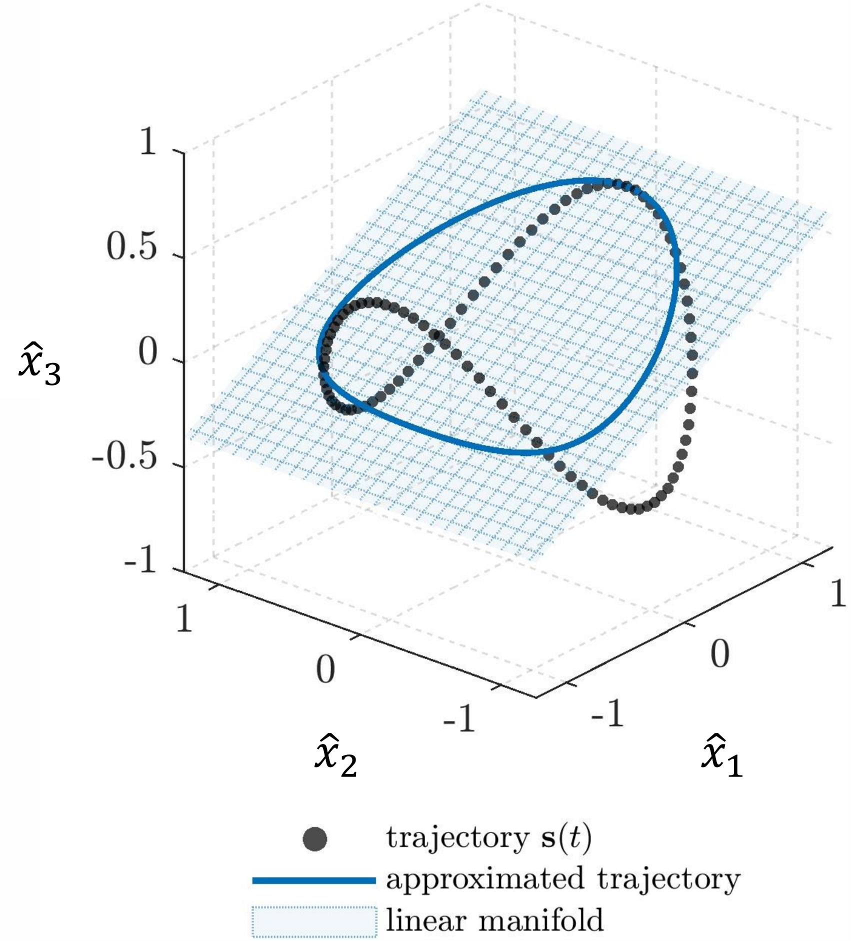
- Generate dataset by uniformly sampling the trajectory at $k = 100$ values of t
- Reference state $\mathbf{x}_{\text{ref}}$ chosen as the initial condition $\mathbf{x}(0)$
- Compute left singular vectors of $\mathbf{X} - \mathbf{X}_{\text{ref}}$:

$$\mathbf{v}_1 = \begin{pmatrix} -0.9347 \\ 0 \\ -0.3554 \end{pmatrix}; \ \mathbf{v}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}; \ \mathbf{v}_3 = \begin{pmatrix} 0.3554 \\ 0 \\ -0.9347 \end{pmatrix}$$

| $\sigma_1 = 12.94$ | $\sigma_2 = 7.04$ | $\sigma_3 = 4.28$ |

Linear approximation with $r = 2$

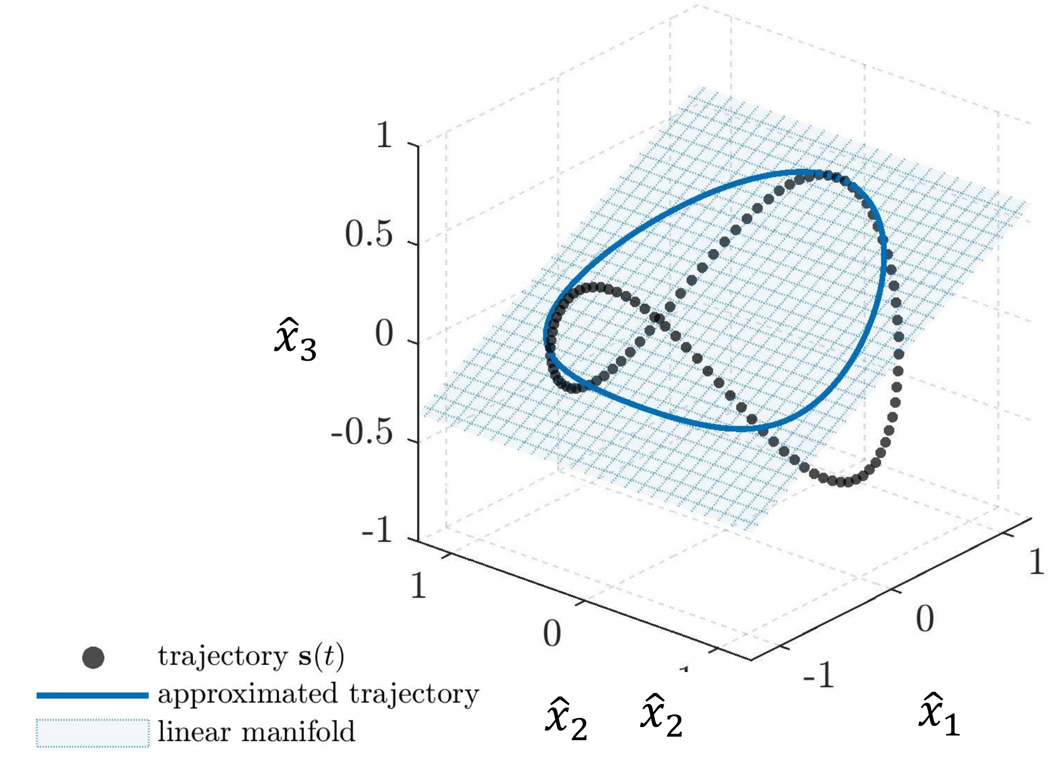
$$\mathbf{x}\,(t) \approx \mathbf{x}_{\text{ref}} + \mathbf{v}_1 \hat{x}_1(t) + \mathbf{v}_2 \hat{x}_2(t)$$



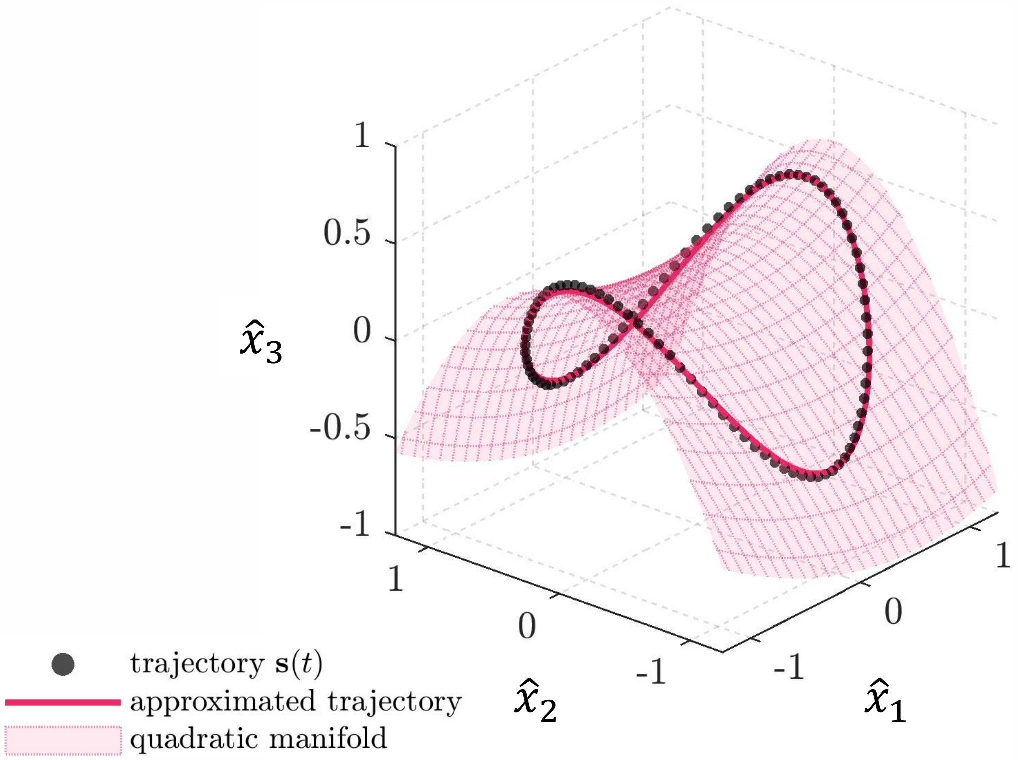
Nonlinear Dimensionality Reduction
Simple illustrative example #1

Choosing $r = 2$, $p = 2$ (quadratic embeddings) we infer from the data the scalar coefficients $(\Xi_1, \Xi_2) = (-0.192, 0.887)$

$$\mathbf{x} \approx \mathbf{x}_{\text{ref}} + \mathbf{v}_1 \hat{x}_1 + \mathbf{v}_2 \hat{x}_2$$



$$\mathbf{x} \approx \mathbf{x}_{\text{ref}} + \mathbf{v}_1 \hat{x}_1 + \mathbf{v}_2 \hat{x}_2 + \mathbf{v}_3 (\Xi_1 \hat{x}_1^2 + \Xi_2 \hat{x}_2^2)$$



Both approximations are of the same dimensionality $r = 2$ (by means of POD coordinates $\hat{x}_1, \hat{x}_2$)

Projecting a linear system

Full-order model (FOM)
state $\mathbf{x} \in \mathbb{R}^N$
input $\mathbf{u} \in \mathbb{R}^{N_i}$

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u}$$

Approximate
$\mathbf{x} \approx \mathbf{V}\hat{\mathbf{x}}$
$\mathbf{V} \in \mathbb{R}^{N \times r}$

Approximation in a linear subspace

Residual: N eqs $\gg$ r dof

$$\mathbf{r} = \mathbf{V}\dot{\hat{\mathbf{x}}} - \mathbf{A}\mathbf{V}\hat{\mathbf{x}} - \mathbf{B}\mathbf{u}$$

Project
$\mathbf{V}^\top \mathbf{r} = 0$

$$\dot{\hat{\mathbf{x}}} = \hat{\mathbf{A}}\hat{\mathbf{x}} + \hat{\mathbf{B}}\mathbf{u}$$

Reduced-order model (ROM)
state $\hat{\mathbf{x}} \in \mathbb{R}^r$
$r \ll N$

$$\hat{\mathbf{A}} = \mathbf{V}^\top \mathbf{A}\mathbf{V}$$
$$\hat{\mathbf{B}} = \mathbf{V}^\top \mathbf{B}$$

Full-order model (FOM)
state $\mathbf{x} \in \mathbb{R}^N$
input $\mathbf{u} \in \mathbb{R}^{N_i}$

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u}$$

Approximate

$$\mathbf{x} \approx \mathbf{V}\hat{\mathbf{x}} + \overline{\mathbf{V}}\boldsymbol{\Xi}\mathbf{g}\big(\hat{\mathbf{x}}(t)\big)$$

$\mathbf{V} \in \mathbb{R}^{N\times r}, \overline{\mathbf{V}} \in \mathbb{R}^{N\times q}, \mathbf{V}^\top\overline{\mathbf{V}}=\mathbf{0}$

Residual: N eqs $\gg r$ dof

$$\mathbf{r} = \mathbf{V}\dot{\hat{\mathbf{x}}} + \overline{\mathbf{V}}\boldsymbol{\Xi}\dot{\mathbf{g}}(\hat{\mathbf{x}}) - \mathbf{A}\mathbf{V}\hat{\mathbf{x}} - \mathbf{A}\overline{\mathbf{V}}\boldsymbol{\Xi}\mathbf{g}\big(\hat{\mathbf{x}}(t)\big) - \mathbf{B}\mathbf{u}$$

Project
$\mathbf{V}^\top\mathbf{r} = 0$

$$\dot{\hat{\mathbf{x}}} = \widehat{\mathbf{A}}\hat{\mathbf{x}} + \widehat{\mathbf{P}}\mathbf{g}\big(\hat{\mathbf{x}}(t)\big) + \widehat{\mathbf{B}}\mathbf{u}$$

Reduced-order model (ROM)
state $\hat{\mathbf{x}} \in \mathbb{R}^r$
$r \ll N$

Projecting a linear system

Approximation in a polynomial manifold

$$\widehat{\mathbf{A}} = \mathbf{V}^\top\mathbf{A}\,\mathbf{V}$$
$$\widehat{\mathbf{P}} = \mathbf{V}^\top\mathbf{A}\overline{\mathbf{V}}\boldsymbol{\Xi}$$
$$\widehat{\mathbf{B}} = \mathbf{V}^\top\mathbf{B}$$

Linear Subspace ROM

FOM: $$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u}$$

$$\mathbf{x} \approx \mathbf{x}_{\mathrm{ref}} + \mathbf{V}\hat{\mathbf{x}}$$

ROM: $$\dot{\hat{\mathbf{x}}} = \hat{\mathbf{c}} + \hat{\mathbf{A}}\hat{\mathbf{x}} + \hat{\mathbf{B}}\mathbf{u}$$

Polynomial Manifold ROM

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u}$$

$$\mathbf{x} \approx \mathbf{x}_{\mathrm{ref}} + \mathbf{V}\hat{\mathbf{x}} + \bar{\mathbf{V}}\mathbf{\Xi}\mathbf{g}(\hat{\mathbf{x}})$$

$$\dot{\hat{\mathbf{x}}} = \hat{\mathbf{c}} + \hat{\mathbf{A}}\hat{\mathbf{x}} + \underbrace{\hat{\mathbf{P}}\mathbf{g}(\hat{\mathbf{x}})} + \hat{\mathbf{B}}\mathbf{u}$$

closure term appears in ROM; representing effects of neglected POD modes

The ROM form is defined by projection theory, but ROM operators $\hat{\mathbf{c}}, \hat{\mathbf{A}}, \hat{\mathbf{H}}, \hat{\mathbf{B}}$ are learned directly from data using Operator Inference

OPERATOR INFERENCE learns the reduced model representation on the polynomial manifold

Define the **structure of the reduced model**

$$\dot{\hat{\mathbf{x}}} = \hat{\mathbf{c}} + \hat{\mathbf{A}}\hat{\mathbf{x}} + \hat{\mathbf{P}}\mathbf{g}(\hat{\mathbf{x}}) + \hat{\mathbf{B}}\mathbf{u}$$

→ **Non-intrusive learning** by inferring reduced operators from simulation data [Peherstorfer & W., 2016]

low-dimensional operators define the reduced model as a dynamical system

snapshots generated from projected simulation data

polynomial terms computed from snapshot data

$$\min_{\hat{\mathbf{c}},\hat{\mathbf{A}},\hat{\mathbf{B}},\hat{\mathbf{P}}} \left\| \hat{\mathbf{c}}^\top + \hat{\mathbf{X}}^\top\hat{\mathbf{A}}^\top + \hat{\mathbf{W}}^\top\hat{\mathbf{P}}^\top + \mathbf{U}^\top\hat{\mathbf{B}}^\top - \dot{\hat{\mathbf{X}}}^\top \right\|$$

minimum residual formulation leads to linear least squares

- All steps of Operator Inference are non-intrusive and highly scalable
- Regularization is often needed to prevent overfitting to training data [McQuarrie, Huang & W., 2021]

Summary: Nonlinear manifold reduced order modeling via Operator Inference

Given a physics model of known form and snapshot data $\mathbf{X} = \begin{pmatrix} | & | & & | \\ \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_k \\ | & | & & | \end{pmatrix}$

1 **Solve the representation learning problem** to compute the nonlinear manifold defined by $\mathbf{V}, \overline{\mathbf{V}}, \boldsymbol{\Xi}, \hat{\mathbf{x}}$ (POD-based approach or Alternating Minimization approach)

2 **Postulate the form of the reduced model** by (pencil and paper) projecting the governing equations onto the nonlinear manifold

3 **Infer the reduced model operators** from projected snapshot data using Operator Inference

Example: Incompressible Navier-Stokes

- Consider the 2D transient flow past a circular cylinder

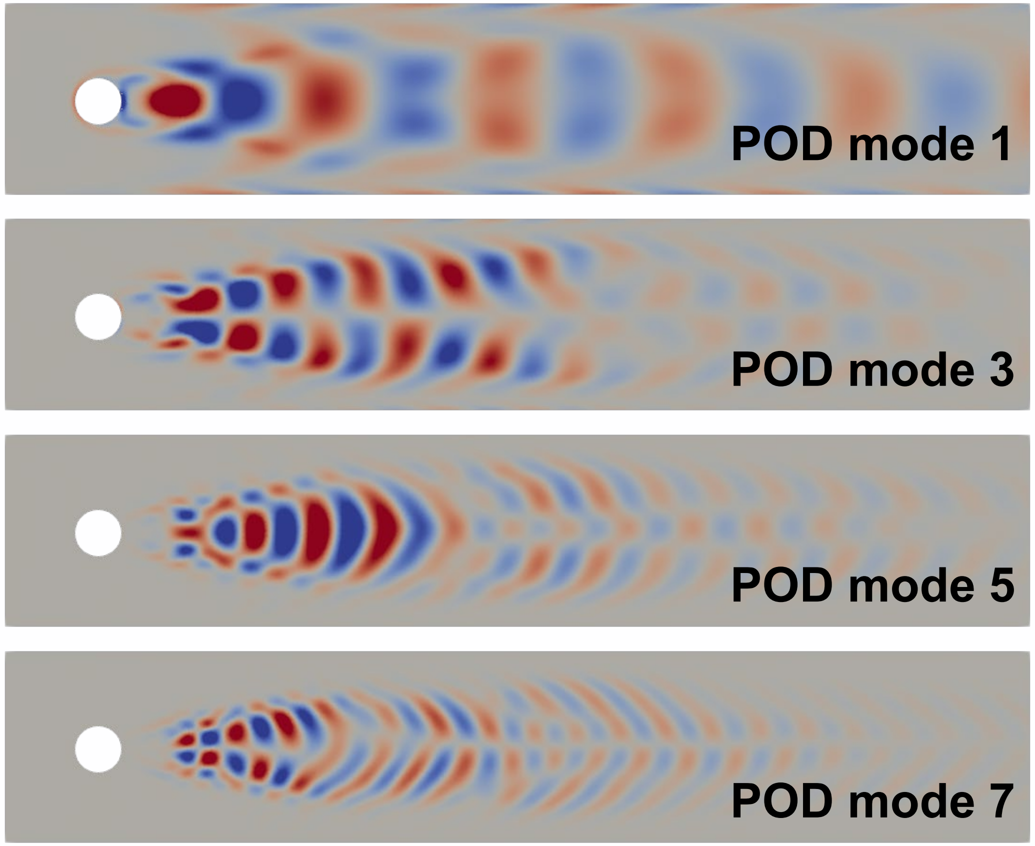
- Governing equations

$$\frac{\partial \mathbf{u}}{\partial t} + \nabla \cdot (\mathbf{u} \otimes \mathbf{u}) = \nabla p + \frac{1}{Re} \Delta \mathbf{u}$$

$$\nabla \cdot \mathbf{u} = 0$$

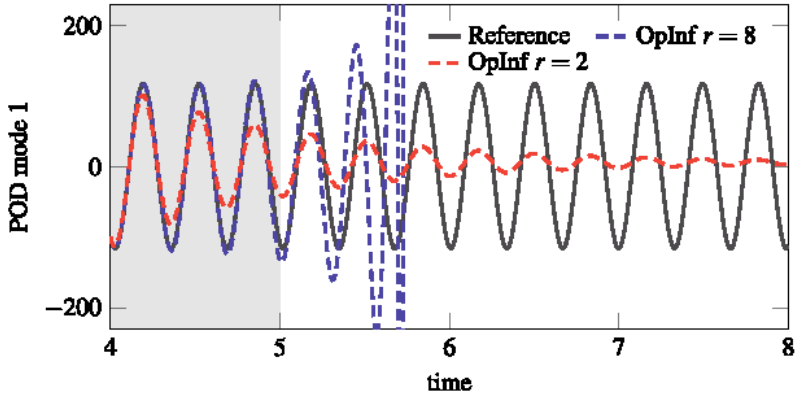
- Because of the problem symmetry, the POD modes come in pair with alternating symmetry properties $(\mathbf{v}_1, \mathbf{v}_2); (\mathbf{v}_3, \mathbf{v}_4); (\mathbf{v}_5, \mathbf{v}_6); (\mathbf{v}_7, \mathbf{v}_8)$

- 8 POD modes capture 99.89% of snapshot energy

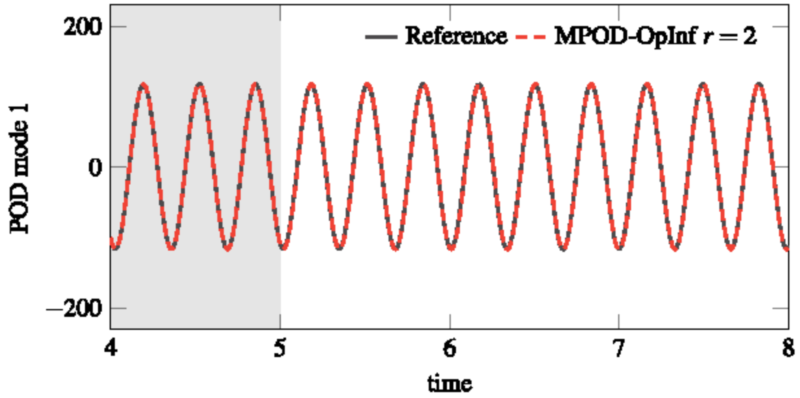


Example: Incompressible Navier-Stokes

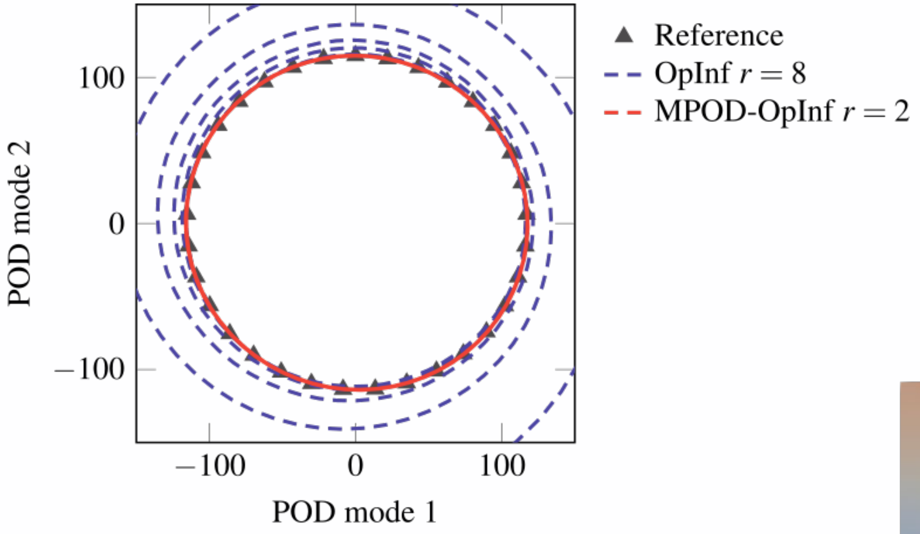
- Linear subspace Operator Inference models and $r = 2$ and $r = 8$
- Nonlinear approximation with quadratic embeddings ($p = 2$)



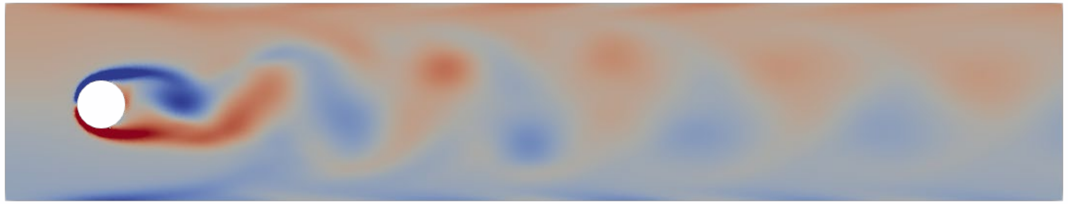
linear subspace models



quadratic manifold models



Stable limit cycle behavior



Flow field prediction (vorticity) at $t = 8$

Reduced-order models are critical enablers for design, control, UQ & digital twins

- Projection-based reduced models preserve physics structure by construction, are interpretable, and are highly effective for a broad range of physics
- Non-intrusive ROM methods exploit the complementary strengths of outside-in and inside-out approaches to surrogate modeling
- Cost of generating ROM training data remains a barrier for many applications → this is a significant research need

Interdisciplinary research & education in computational engineering & sciences

advancing computational science to address society's grand challenges

ODEN.UTEXAS.EDU

ODEN INSTITUTE
FOR COMPUTATIONAL ENGINEERING & SCIENCES