MAX PLANCK INSTITUTE
FOR DYNAMICS OF COMPLEX
TECHNICAL SYSTEMS
MAGDEBURG

CSC
COMPUTATIONAL METHODS IN
SYSTEMS AND CONTROL THEORY

LQResNet: Using DNNs for Learning of Dynamical Systems

Peter Benner

Joint work with Pawan Goyal (and others. . .)

Workshop on Control of Dynamical Systems
14–16 June 2021, Dubrovnik, Croatia

Supported by:

DFG-Graduiertenkolleg
MATHEMATISCHE
KOMPLEXITÄTSREDUKTION

BiG max MAX PLANCK RESEARCH NETWORK on big-data-driven materials science

hrzz Hrvatska zaklada za znanost

ConDyS

Outline

Joint Work With . . .

Pawan Goyal
MPI Magdeburg

Karen Willcox
Oden Institute,
UT Austin

Benjamin Peherstorfer
Courant Institute, NYU

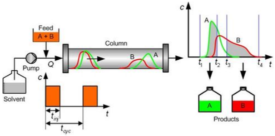
Boris Kramer
UCSD

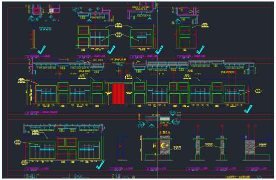
Igor Pontes Duff
MPI Magdeburg

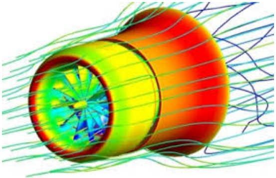
Jan Heiland
MPI Magdeburg

Bülent Karasözen
METU, Ankara

Süleyman Yıldız
METU, Ankara







Dynamic models are important

- to analyze transient behavior under operating conditions;
- for controller design;
- design studies w.r.t. (material/geometry) parameter variations;
- long-time horizon reliability prediction.

Problem set-up

- Construct a mathematical model

$$\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t))$$

describing the dynamics of the process.

Problem set-up

- Construct a mathematical model

$$\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t))$$

  describing the dynamics of the process.
- Neural network-based approaches: e.g., recurrent neural networks and long short time memory networks.

Learning Dynamics from Data

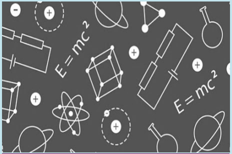
Problem set-up

- Construct a mathematical model

$$\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t))$$

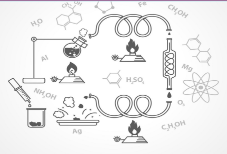
  describing the dynamics of the process.
- Neural network-based approaches: e.g., recurrent neural networks and long short time memory networks.
- Leverage all prior information about the process for efficient learning.

Key sources of information



Physical laws

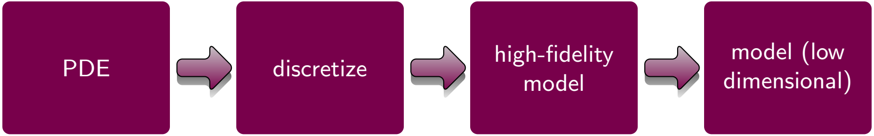
Domain knowledge



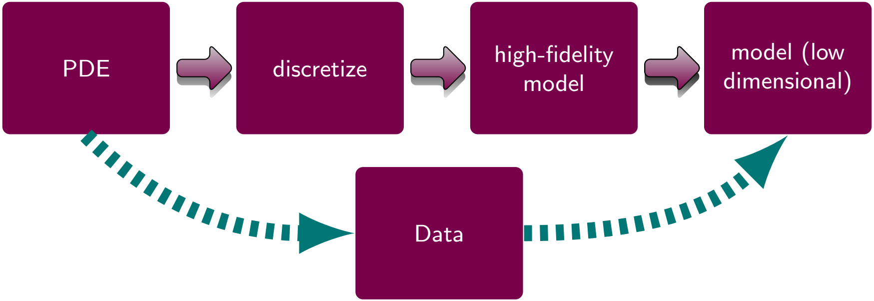
Collected data

Learning Dynamics from Data

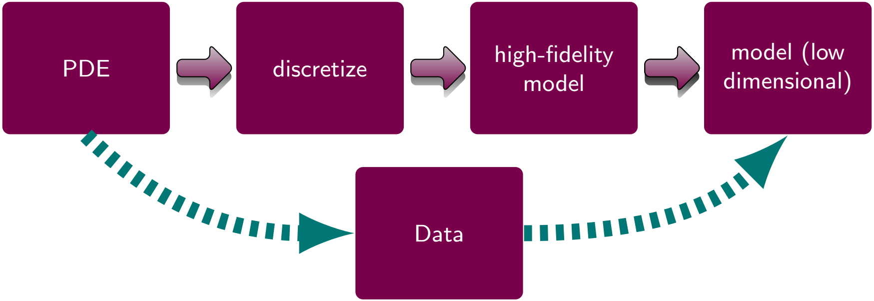
- Engineering processes are supported by domain knowledge and first principles
  $\rightsquigarrow$ a PDE model can be obtained that adequately explains the dynamics



- Engineering processes are supported by domain knowledge and first principles
  $\rightsquigarrow$ a PDE model can be obtained that adequately explains the dynamics

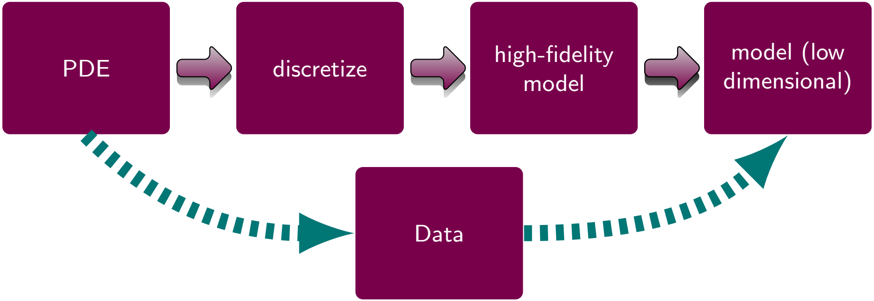


- Engineering processes are supported by domain knowledge and first principles
  $\rightsquigarrow$ a PDE model can be obtained that adequately explains the dynamics



- Data collection: obtained using a legacy code, or commercial software, or experiments.

- Engineering processes are supported by domain knowledge and first principles
  $\rightsquigarrow$ a PDE model can be obtained that adequately explains the dynamics



- Data collection: obtained using a legacy code, or commercial software, or experiments.
- Ideal goal: obtain the same reduced-order model (ROM) as obtained by intrusive model order reduction using data, so that error bounds and convergence analysis for ROMs can be directly employed!

Operator inference framework [Peherstorfer/Willcox '16]

- Operator inference leverages the known physical structure at the PDE level.
- Assume a quadratic high-fidelity model resulting from an underlying PDE $\frac{\partial x}{\partial t} = \mathcal{A}(x) + \mathcal{H}(x)$ with linear and quadratic terms:

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

Operator inference framework [Peherstorfer/Willcox '16]

- Operator inference leverages the known physical structure at the PDE level.
- Assume a quadratic high-fidelity model resulting from an underlying PDE $\frac{\partial x}{\partial t} = \mathcal{A}(x) + \mathcal{H}(x)$ with linear and quadratic terms:

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

- Data preparation (in reduced dimension)

  1. Build temporal snapshot matrix $\mathbf{X} := \begin{bmatrix} | & | & & | \\ \mathbf{x}_0 & \mathbf{x}_1 & \cdots & \mathbf{x}_k \\ | & | & & | \end{bmatrix}$.

Operator inference framework

[Peherstorfer/Willcox '16]

- Operator inference leverages the known physical structure at the PDE level.
- Assume a quadratic high-fidelity model resulting from an underlying PDE $\frac{\partial x}{\partial t} = \mathcal{A}(x) + \mathcal{H}(x)$ with linear and quadratic terms:

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

- Data preparation (in reduced dimension)

1. Build temporal snapshot matrix $\mathbf{X} := \begin{bmatrix} | & | & & | \\ \mathbf{x}_0 & \mathbf{x}_1 & \cdots & \mathbf{x}_k \\ | & | & & | \end{bmatrix}$.
2. Compute projection matrix $\mathbf{V}$ using dominant POD basis vectors.

Operator inference framework [Peherstorfer/Willcox '16]

- Operator inference leverages the known physical structure at the PDE level.
- Assume a quadratic high-fidelity model resulting from an underlying PDE $\frac{\partial x}{\partial t} = \mathcal{A}(x) + \mathcal{H}(x)$ with linear and quadratic terms:

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

- Data preparation (in reduced dimension)

1. Build temporal snapshot matrix $\mathbf{X} := \begin{bmatrix} | & | & & | \\ \mathbf{x}_0 & \mathbf{x}_1 & \cdots & \mathbf{x}_k \\ | & | & & | \end{bmatrix}$.
2. Compute projection matrix $\mathbf{V}$ using dominant POD basis vectors.
3. Reduced state vectors

$$\hat{\mathbf{X}} := V^T\mathbf{X} = \begin{bmatrix} | & | & & | \\ \hat{\mathbf{x}}_0 & \hat{\mathbf{x}}_1 & \cdots & \hat{\mathbf{x}}_k \\ | & | & & | \end{bmatrix}, \qquad \hat{\mathbf{X}}^{\otimes} := \begin{bmatrix} | & | & & | \\ \hat{\mathbf{x}}_0^{\otimes} & \hat{\mathbf{x}}_1^{\otimes} & \cdots & \hat{\mathbf{x}}_k^{\otimes} \\ | & | & & | \end{bmatrix}.$$

with $\hat{\mathbf{x}}_i = \mathbf{V}^\top \mathbf{x}_i$ and $\hat{\mathbf{x}}_i^{\otimes} = \hat{\mathbf{x}}_i \otimes \hat{\mathbf{x}}_i$

Operator inference framework

[Peherstorfer/Willcox '16]

- Operator inference leverages the known physical structure at the PDE level.
- Assume a quadratic high-fidelity model resulting from an underlying PDE $\frac{\partial x}{\partial t} = \mathcal{A}(x) + \mathcal{H}(x)$ with linear and quadratic terms:

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

- Data preparation (in reduced dimension)

1. Build temporal snapshot matrix $\mathbf{X} := \begin{bmatrix} | & | & & | \\ \mathbf{x}_0 & \mathbf{x}_1 & \cdots & \mathbf{x}_k \\ | & | & & | \end{bmatrix}$.
2. Compute projection matrix $\mathbf{V}$ using dominant POD basis vectors.
3. Reduced state vectors

$$\hat{\mathbf{X}} := V^T\mathbf{X} = \begin{bmatrix} | & | & & | \\ \hat{\mathbf{x}}_0 & \hat{\mathbf{x}}_1 & \cdots & \hat{\mathbf{x}}_k \\ | & | & & | \end{bmatrix}, \qquad \hat{\mathbf{X}}^{\otimes} := \begin{bmatrix} | & | & & | \\ \hat{\mathbf{x}}_0^{\otimes} & \hat{\mathbf{x}}_1^{\otimes} & \cdots & \hat{\mathbf{x}}_k^{\otimes} \\ | & | & & | \end{bmatrix}.$$

4. Approximate time-derivative data $\dot{\hat{\mathbf{X}}} := \begin{bmatrix} | & | & & | \\ \dot{\hat{\mathbf{x}}}_0 & \dot{\hat{\mathbf{x}}}_1 & \cdots & \dot{\hat{\mathbf{x}}}_k \\ | & | & & | \end{bmatrix}$.

Operator inference framework

[Peherstorfer/Willcox '16]

A ROM of the form

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

can be obtained using projected data by solving the optimization problem

$$\min_{\hat{\mathbf{A}},\hat{\mathbf{H}}} \left\| \dot{\hat{\mathbf{X}}} - \hat{\mathbf{A}}\hat{\mathbf{X}} - \hat{\mathbf{H}}\hat{\mathbf{X}}^{\otimes} \right\|.$$

Operator inference framework

[Peherstorfer/Willcox '16]

A ROM of the form

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

can be obtained using projected data by solving the optimization problem

$$\min_{\hat{\mathbf{A}},\hat{\mathbf{H}}} \left\| \dot{\hat{\mathbf{X}}} - \hat{\mathbf{A}}\hat{\mathbf{X}} - \hat{\mathbf{H}}\hat{\mathbf{X}}^{\otimes} \right\|.$$

Remarks:

- Notice that we do not require at any step the full-order discretized model.

Operator inference framework [Peherstorfer/Willcox '16]

A ROM of the form

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

can be obtained using projected data by solving the optimization problem

$$\min_{\hat{\mathbf{A}},\hat{\mathbf{H}}} \left\| \dot{\hat{\mathbf{X}}} - \hat{\mathbf{A}}\hat{\mathbf{X}} - \hat{\mathbf{H}}\hat{\mathbf{X}}^{\otimes} \right\|.$$

Remarks:

- Notice that we do not require at any step the full-order discretized model.
- Operator inference recovers intrusive POD reduced model if data are Markovian.

[Peherstorfer '20]

Operator inference framework [Peherstorfer/Willcox '16]

A ROM of the form

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

can be obtained using projected data by solving the optimization problem

$$\min_{\hat{\mathbf{A}},\hat{\mathbf{H}}} \left\| \dot{\hat{\mathbf{X}}} - \hat{\mathbf{A}}\hat{\mathbf{X}} - \hat{\mathbf{H}}\hat{\mathbf{X}}^{\otimes} \right\| + \mathcal{R}(\hat{\mathbf{A}}, \hat{\mathbf{H}}).$$

Remarks:

- Notice that we do not require at any step the full-order discretized model.
- Operator inference recovers intrusive POD reduced model if data are Markovian.
  [Peherstorfer '20]
- Typically, the least-squares problem is ill-conditioned, hence need regularization.
  [McQuarrie et al. '21, B./Goyal/Heiland/Pontes '21]

Nonlinear systems [B./Goyal/Kramer/Peherstorfer/Willcox '20]

- Consider a nonlinear system of the form

$$\frac{\partial s}{\partial t} = \mathcal{A}(s) + \mathcal{H}(s) + \mathcal{F}(t, s),$$

where the analytic form of $\mathcal{F}(t, s)$ is known.

- We can learn a ROM of the form

$$\dot{\hat{\mathbf{s}}}(t) = \hat{\mathbf{A}}\hat{\mathbf{s}} + \hat{\mathbf{H}}\left(\hat{\mathbf{s}} \otimes \hat{\mathbf{s}}\right) + \hat{\mathbf{f}}(t, \hat{\mathbf{s}})$$

directly from data!

- Simulation of reduced nonlinear system can be further accelerated using hyper-reduction.

- The dynamics of a batch chromatography column can be described by the coupled PDE system of advection-diffusion type:

$$\frac{\partial c_i}{\partial t} + \frac{1-\epsilon}{\epsilon}\frac{\partial q_i}{\partial t} + \frac{\partial c_i}{\partial x} - \frac{1}{\mathrm{Pe}}\frac{\partial^2 c_i}{\partial x^2} = 0,$$
$$\frac{\partial q_i}{\partial t} = \kappa_i \left(q_i^{Eq} - q_i \right).$$

- It is a coupled PDE; thus, the coupling structure is desired to be preserved in learned ROM
- This is achieved by block diagonal projection, thereby not mixing separate physical quantities.

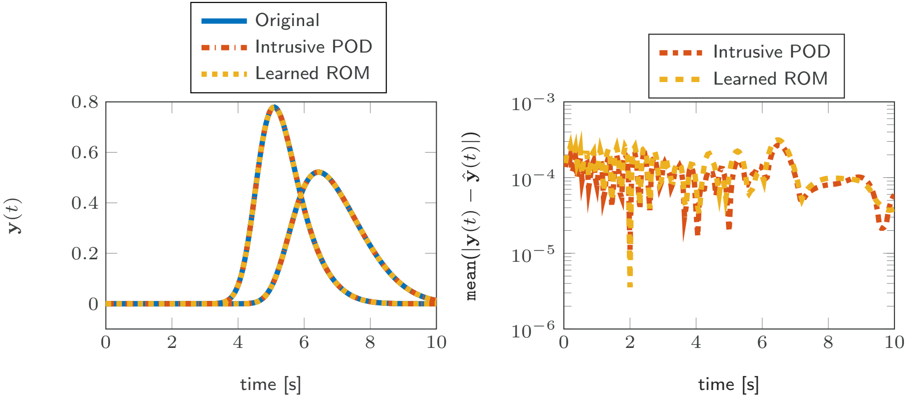


Figure: Batch chromatography example: A comparison of the POD intrusive model with the learned model of order $r = 4 \times 22$, where $n = 1600$ and $\mathrm{Pe} = 2000$.

Parameterized Shallow Water Equations

- Parameterized shallow water equations are given by [Yildiz et al '20]

$$\frac{\partial}{\partial t}\tilde{u} = -h_x + \sin\theta\ \tilde{v} - \tilde{u}\tilde{u}_x - \tilde{v}\tilde{u}_y + \delta\cos\theta(h\tilde{u})_x - \frac{3}{8}\left(\delta\cos\theta\right)^2\left(h^2\right)_x,$$
$$\frac{\partial}{\partial t}\tilde{v} = -h_y + \sin\theta\ \tilde{u} + \frac{1}{2}\delta\sin\theta\cos\theta\ h - \tilde{u}\tilde{v}_x - \tilde{v}\tilde{v}_y$$
$$+\delta\cos\theta\left((h\tilde{u})_y + \frac{1}{2}h\left(\tilde{v}_x - \tilde{u}_y\right)\right) - \frac{3}{8}\left(\delta\cos\theta\right)^2\left(h^2\right)_y,$$
$$\frac{\partial}{\partial t}h = -(h\tilde{u})_x - (h\tilde{v})_y + \frac{1}{2}\delta\cos\theta(h^2)_x.$$

- Parameterized by the latitude θ.
- $\tilde{\mathbf{u}} =: (\tilde{u}; \tilde{v})$ is the canonical velocity.
- h is the height field.
- We collect the training data for 5 different parameter realizations θ in $\left[\frac{\pi}{6}, \frac{\pi}{3}\right]$.
- Infer a reduced parametric model directly from data of order $r = 75$.

Parameterized Shallow Water Equations

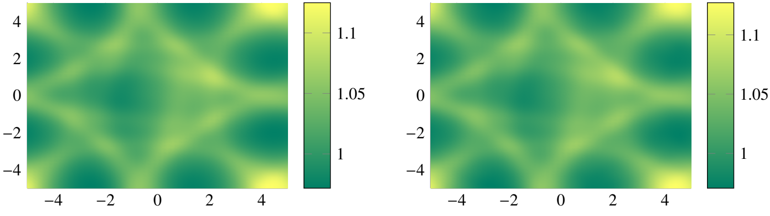
- Parameterized shallow water equations are given by [Yildiz et al '20]

$$\frac{\partial}{\partial t}\tilde{u} = -h_x + \sin\theta\ \tilde{v} - \tilde{u}\tilde{u}_x - \tilde{v}\tilde{u}_y + \delta\cos\theta(h\tilde{u})_x - \frac{3}{8}\left(\delta\cos\theta\right)^2\left(h^2\right)_x,$$

$$\frac{\partial}{\partial t}\tilde{v} = -h_y + \sin\theta\ \tilde{u} + \frac{1}{2}\delta\sin\theta\cos\theta\ h - \tilde{u}\tilde{v}_x - \tilde{v}\tilde{v}_y$$
$$+\,\delta\cos\theta\left((h\tilde{u})_y + \frac{1}{2}h\left(\tilde{v}_x - \tilde{u}_y\right)\right) - \frac{3}{8}\left(\delta\cos\theta\right)^2\left(h^2\right)_y,$$

$$\frac{\partial}{\partial t}h = -(h\tilde{u})_x - (h\tilde{v})_y + \frac{1}{2}\delta\cos\theta(h^2)_x.$$

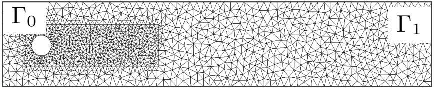
- Comparison of the height field for the parameter $\theta = \frac{5\pi}{24}$:

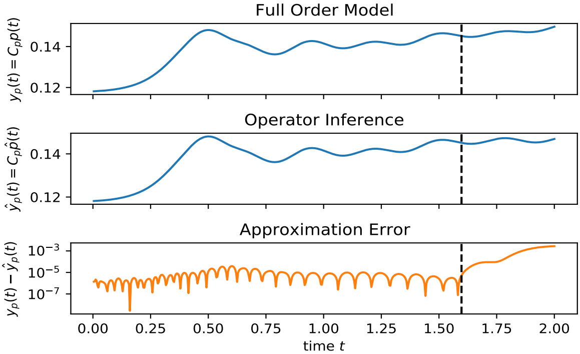


(a) FOM

(b) Learned parametric model

Tailored operator inference for incompressible Navier-Stokes equations, by heeding incompressibility condition. [B./Goyal/Heiland/Pontes '21]





Combining Operator Inference with Deep Learning

Operator Inference for General Nonlinear Systems

Problem formulation

$$\dot{\mathbf{v}}(t) = \mathbf{f}(\mathbf{v}(t)) + \mathbf{r}(\mathbf{v}(t))$$

- $\mathbf{f}(\mathbf{v}(t))$: known from physical laws or expert knowledge;
  - e.g., for chemical reaction models, we expect to have an Arrhenius-type term.
- $\mathbf{r}(\mathbf{v}(t))$: unknown terms
  - e.g., friction terms in robotics or vibration systems, effects of removed higher-frequency dynamics on the low-frequency response, etc.

Operator Inference for General Nonlinear Systems

Problem formulation

$$\dot{\mathbf{v}}(t) = \mathbf{f}(\mathbf{v}(t)) + \mathbf{r}(\mathbf{v}(t))$$

- $\mathbf{f}(\mathbf{v}(t))$: known from physical laws or expert knowledge;
  - e.g., for chemical reaction models, we expect to have an Arrhenius-type term.
- $\mathbf{r}(\mathbf{v}(t))$: unknown terms
  - e.g., friction terms in robotics or vibration systems, effects of removed higher-frequency dynamics on the low-frequency response, etc.

Observation

- Often, governing equations are **quadratic**, i.e.,

  $\mathbf{f}(\mathbf{v}) := \mathbf{A}\mathbf{v} + \mathbf{H}\left(\mathbf{v} \otimes \mathbf{v}\right).$
- If no additional information is given, we assume $\mathbf{f}$ to be quadratic.
- Moreover, possible to find artificial variables in which dynamics are quadratic.

  Philosophy: Lift & learn [Qian et al. '20]

Navier-Stokes equations

$$\rho\left(\frac{\partial u_r}{\partial t} + u_r\frac{\partial u_r}{\partial r} + u_z\frac{\partial u_r}{\partial z}\right) = -\frac{\partial p}{\partial r} + \mu\left(\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial u_r}{\partial r}\right) + \frac{\partial^2 u_r}{\partial z^2} - \frac{u_r}{r^2}\right) + \rho g_r$$

$$\rho\left(\frac{\partial u_z}{\partial t} + u_r\frac{\partial u_z}{\partial r} + u_z\frac{\partial u_z}{\partial z}\right) = -\frac{\partial p}{\partial z} + \mu\left(\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial u_z}{\partial r}\right) + \frac{\partial^2 u_z}{\partial z^2}\right) + \rho g_z$$

$$\frac{1}{r}\frac{\partial}{\partial r}(r u_r) + \dots$$

Fisher's equation

$$u_t = u(1-u) + u_{xx}$$

Problem formulation

$$\dot{\mathbf{v}}(t) = \mathbf{f}(\mathbf{v}(t)) + \mathbf{r}(\mathbf{v}(t))$$

- $\mathbf{f}(\mathbf{v}(t))$: known from physical laws or expert knowledge;
  - e.g., for chemical reaction models, we expect to have an Arrhenius-type term.
- $\mathbf{r}(\mathbf{v}(t))$: unknown terms
  - e.g., friction terms in robotics or vibration systems, effects of removed higher-frequency dynamics on the low-frequency response, etc.

Lifting

[Gu '09/'11, Benner/Breiten '15, Qian et al '20]

Consider the nonlinear system:

$$\dot{\mathbf{x}} = -\mathbf{x} + e^{-\mathbf{x}}$$

Problem formulation

$$\dot{\mathbf{v}}(t) = \mathbf{f}(\mathbf{v}(t)) + \mathbf{r}(\mathbf{v}(t))$$

- $\mathbf{f}(\mathbf{v}(t))$: known from physical laws or expert knowledge;
  - e.g., for chemical reaction models, we expect to have an Arrhenius-type term.
- $\mathbf{r}(\mathbf{v}(t))$: unknown terms
  - e.g., friction terms in robotics or vibration systems, effects of removed higher-frequency dynamics on the low-frequency response, etc.

Lifting

[Gu '09/'11, Benner/Breiten '15, Qian et al '20]

Consider the nonlinear system:

$$\dot{\mathbf{x}} = -\mathbf{x} + e^{-\mathbf{x}}$$

- Define $\mathbf{z}(t) = e^{-\mathbf{x}} \rightsquigarrow \dot{\mathbf{z}}(t) = -e^{-\mathbf{x}}\dot{\mathbf{x}} = -\mathbf{z}(t)\left(-\mathbf{x}(t) + \mathbf{z}(t)\right)$

Operator Inference for General Nonlinear Systems

Problem formulation

$$\dot{\mathbf{v}}(t) = \mathbf{f}(\mathbf{v}(t)) + \mathbf{r}(\mathbf{v}(t))$$

- $\mathbf{f}(\mathbf{v}(t))$: known from physical laws or expert knowledge;
  - e.g., for chemical reaction models, we expect to have an Arrhenius-type term.
- $\mathbf{r}(\mathbf{v}(t))$: unknown terms
  - e.g., friction terms in robotics or vibration systems, effects of removed higher-frequency dynamics on the low-frequency response, etc.

Lifting [Gu '09/'11, Benner/Breiten '15, Qian et al '20]

Consider the nonlinear system:

$$\dot{\mathbf{x}} = -\mathbf{x} + e^{-\mathbf{x}}$$

- Define $\mathbf{z}(t) = e^{-\mathbf{x}} \rightsquigarrow \dot{\mathbf{z}}(t) = -e^{-\mathbf{x}}\dot{\mathbf{x}} = -\mathbf{z}(t)\left(-\mathbf{x}(t) + \mathbf{z}(t)\right)$
- The system becomes linear-quadratic in $(\mathbf{x}(t), \mathbf{z}(t))$, i.e.,

$$\begin{bmatrix} \dot{\mathbf{x}}(t) \\ \dot{\mathbf{z}}(t) \end{bmatrix} = \begin{bmatrix} -\mathbf{x}(t) + z(t) \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ \mathbf{z}(t)\left(\mathbf{x}(t) - \mathbf{z}(t)\right) \end{bmatrix}.$$

Linear-Quadratic Residual Networks

For simplicity, consider the form:

$$\dot{\mathbf{v}}(t) = f(\mathbf{v}(t)) = \mathbf{A}\mathbf{v}(t) + \mathbf{H}\left(\mathbf{v}(t) \otimes \mathbf{v}(t)\right) + \mathbf{r}(\mathbf{v}(t)),$$

where

- $\mathbf{r}(\mathbf{v}(t))$ can be interpreted as a residual that cannot be resolved by the quadratic-form or prior knowledge.

For simplicity, consider the form:

$$\dot{\mathbf{v}}(t) = f(\mathbf{v}(t)) = \mathbf{A}\mathbf{v}(t) + \mathbf{H}\left(\mathbf{v}(t) \otimes \mathbf{v}(t)\right) + \mathbf{r}(\mathbf{v}(t)),$$

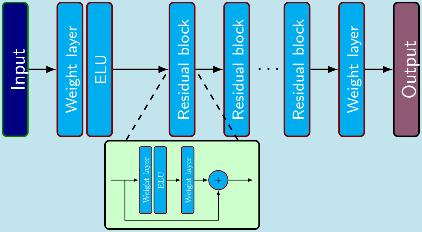
where

- $\mathbf{r}(\mathbf{v}(t))$ can be interpreted as a residual that cannot be resolved by the quadratic-form or prior knowledge.

Residual networks

[He/Ren/Sun '16]

- Have shown their power in computer vision applications.



For simplicity, consider the form:

$$\dot{\mathbf{v}}(t) = f(\mathbf{v}(t)) = \mathbf{A}\mathbf{v}(t) + \mathbf{H}\left(\mathbf{v}(t) \otimes \mathbf{v}(t)\right) + \mathbf{r}(\mathbf{v}(t)),$$

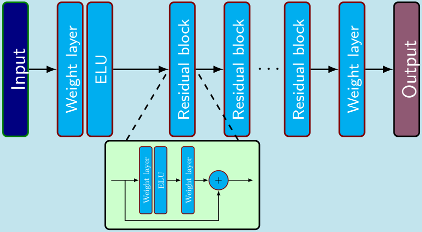
where

- $\mathbf{r}(\mathbf{v}(t))$ can be interpreted as a residual that cannot be resolved by the quadratic-form or prior knowledge.

Residual networks

[He/Ren/Sun '16]

- Have shown their power in computer vision applications.
- There is an established connection to dynamical systems.



For simplicity, consider the form:

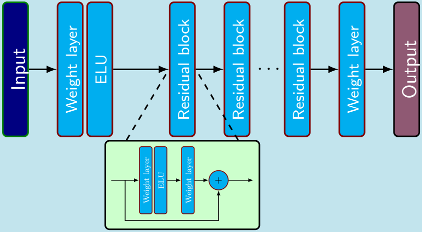
$$\dot{\mathbf{v}}(t) = f(\mathbf{v}(t)) = \mathbf{A}\mathbf{v}(t) + \mathbf{H}\left(\mathbf{v}(t) \otimes \mathbf{v}(t)\right) + \mathbf{r}(\mathbf{v}(t)),$$

where

- $\mathbf{r}(\mathbf{v}(t))$ can be interpreted as a residual that cannot be resolved by the quadratic-form or prior knowledge.

Residual networks [He/Ren/Sun '16]

- Have shown their power in computer vision applications.
- There is an established connection to dynamical systems.
- Residual type connections hint to adaptive refinement of solution or features.



Linear-Quadratic Residual Networks

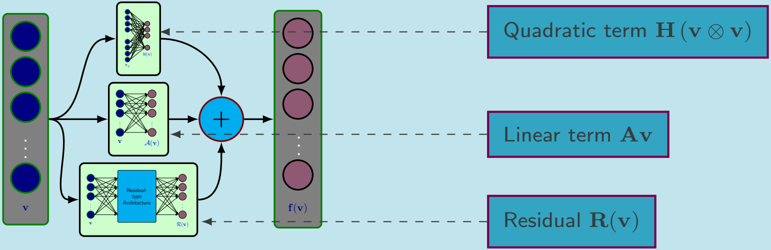
For simplicity, consider the form:

$$\dot{\mathbf{v}}(t) = f(\mathbf{v}(t)) = \mathbf{A}\mathbf{v}(t) + \mathbf{H}\left(\mathbf{v}(t) \otimes \mathbf{v}(t)\right) + \mathbf{r}(\mathbf{v}(t)),$$

where

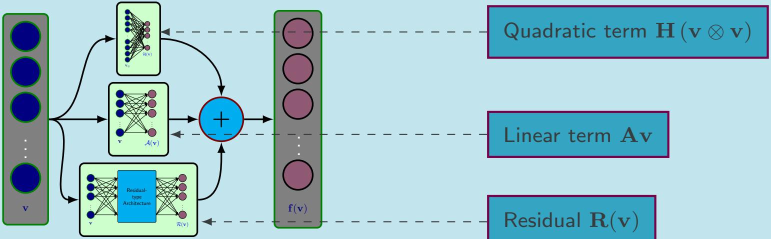
- $\mathbf{r}(\mathbf{v}(t))$ can be interpreted as a residual that cannot be resolved by the quadratic-form or prior knowledge.

Linear-Quadratic Residual Networks (LQResNet) [Goyal/B. '21]



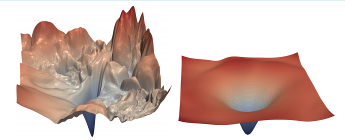
Advantages of the Architecture

Linear-Quadratic Residual Networks (LQResNet) [Goyal/B. '21]



Remarks

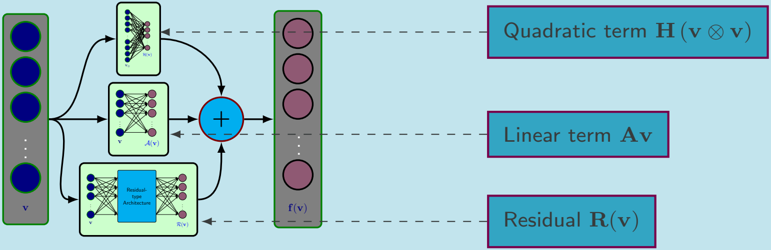
- Due to skip connections, loss landscape becomes less bumpy. [Li et al. '18]



(a) without skip connections (b) with skip connections

Advantages of the Architecture

Linear-Quadratic Residual Networks (LQResNet) [Goyal/B. '21]

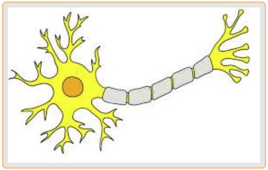


Remarks

- Due to skip connections, loss landscape becomes less bumpy. [Li et al. '18]
- Layers can be added without restarting whole optimization as deep residual layers tend to refine the mapping.

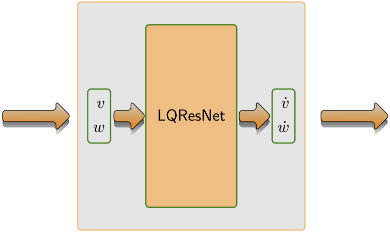
Set-up

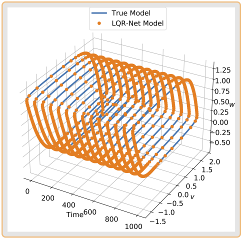
- The FitzHugh-Nagumo model is a coupled PDE-ODE model describing the spiking of a neuron.
- Assume to have time-series data for 10 different initial conditions.
- We build different networks for both variables.
- We check the predictive capabilities of the inferred model under new initial condition.



Governing equations

$$\dot{v}(t) = v - \frac{v^3}{3} - w + RI_{ext}$$
$$\dot{w}(t) = v + a - bw$$





Set-up

- Represents complex wide-range dynamical behavior in yeast glycolysis.

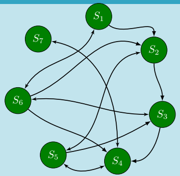


Figure: Interaction topology for 7 species.

Set-up

- Represents complex wide-range dynamical behavior in yeast glycolysis.
- There are 7 involved species.

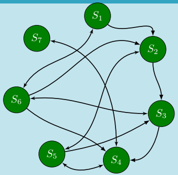


Figure: Interaction topology for 7 species.

Set-up

- Represents complex wide-range dynamical behavior in yeast glycolysis.
- There are 7 involved species.
- Data for 30 different initial conditions.

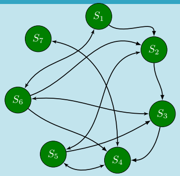


Figure: Interaction topology for 7 species.

Set-up

- Represents complex wide-range dynamical behavior in yeast glycolysis.
- There are 7 involved species.
- Data for 30 different initial conditions.
- Utilized interaction topology in learning.

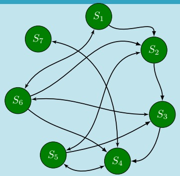


Figure: Interaction topology for 7 species.

Set-up

- Represents complex wide-range dynamical behavior in yeast glycolysis.
- There are 7 involved species.
- Data for 30 different initial conditions.
- Utilized interaction topology in learning.
- Check the predictive capabilities under new condition.

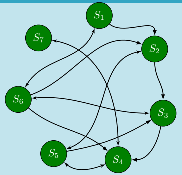
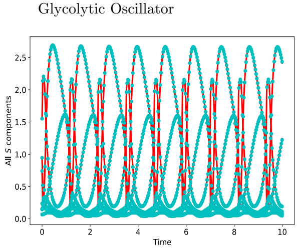
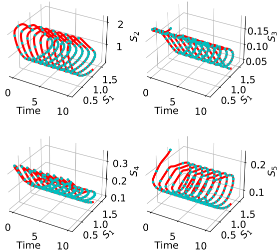


Figure: Interaction topology for 7 species.





Tubular Reactor Model

- One dimensional model with a single reaction, describing dynamics of the species concentration $\psi(x,t)$ and temperature $\theta(x,t)$ via

$$\frac{\partial \psi}{\partial t} = \frac{1}{\mathrm{Pe}}\frac{\partial^2 \psi}{\partial x^2} - \frac{\partial \psi}{\partial x} - \mathcal{D}\mathcal{F}(\psi,\theta;\gamma),$$
$$\frac{\partial \theta}{\partial t} = \frac{1}{\mathrm{Pe}}\frac{\partial^2 \theta}{\partial x^2} - \frac{\partial \theta}{\partial x} - \beta(\theta - \theta_{\text{ref}}) + \mathcal{B}\mathcal{D}\mathcal{F}(\psi,\theta;\gamma),$$

with spatial variable $x \in (0,1)$, time $t > 0$ and Arrhenius reaction term

$$\mathcal{F}(\psi,\theta;\gamma) = \psi \exp\left(\gamma - \frac{\gamma}{\theta}\right).$$

- The quantity of interest is the temperature oscillation at the reactor exit:

$$\mathbf{y}(t) = \theta(\mathbf{x} = 1, t).$$

Tubular Reactor Model

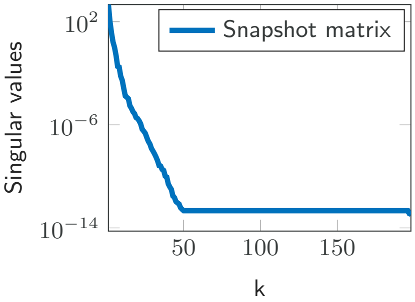


Figure: Decay of singular values of the snapshots.

- Rapid decay of singular values of training data $\rightsquigarrow$ possibility of lower order models.
- The dominant three POD modes capture more than 99.8% of the energy, yet the POD model is unstable.

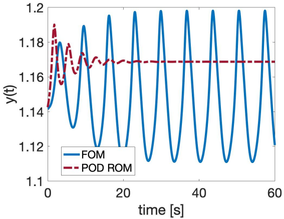


Figure: A comparison of the temperature oscillations at exit.

- Rapid decay of singular values of training data $\rightsquigarrow$ possibility of lower order models.
- The dominant three POD modes capture more than 99.8% of the energy, yet the POD model is unstable.
- We employ the LQResNet approach to learn the correction.

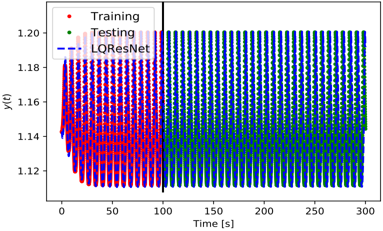


Figure: A comparison of the temperature oscillations at exit.

- Rapid decay of singular values of training data $\rightsquigarrow$ possibility of lower order models.
- The dominant three POD modes capture more than 99.8% of the energy, yet the POD model is unstable.
- We employ the LQResNet approach to learn the correction.

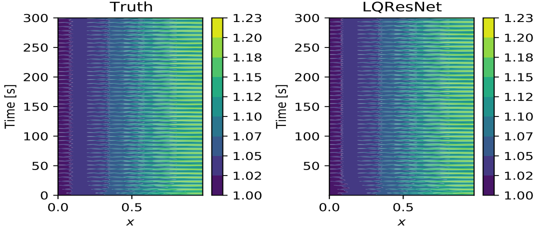


Figure: A comparison of the temperature oscillations in the whole domain.

- Rapid decay of singular values of training data $\rightsquigarrow$ possibility of lower order models.
- The dominant three POD modes capture more than 99.8% of the energy, yet the POD model is unstable.
- We employ the LQResNet approach to learn the correction.

Contribution

- We have studied an approach to learn a mathematical model to describe nonlinear dynamics.
- Basis: operator inference and its extensions, utilizing prior PDE knowledge.
- New: model residual identified using architecture LQResNet, inspired by residual network.
- The design allows us to incorporate prior hypotheses about the process.

Outlook

Contribution

- We have studied an approach to learn a mathematical model to describe nonlinear dynamics.
- Basis: operator inference and its extensions, utilizing prior PDE knowledge.
- New: model residual identified using architecture LQResNet, inspired by residual network.
- The design allows us to incorporate prior hypotheses about the process.

On-going work

- Very often, we can build a dictionary of good candidate basis functions, but probably do not want all of them in the dictionary. Therefore, we seek a parsimonious model
  - to pick few entries from the dictionary and learn residual by deep learning.
- Appropriate treatment of noise ... [Rudy/Kutz/Brunton '19]
- Missing/corrupted data in time series.
- Working with several applications in material science and chemical engineering.

Contribution

- We have studied an approach to learn a mathematical model to describe nonlinear dynamics.
- Basis: operator inference and its extensions, utilizing prior PDE knowledge.
- New: model residual identified using architecture LQResNet, inspired by residual network.
- The design allows us to incorporate prior hypotheses about the process.

On-going work

- Very often, we can build a dictionary of good candidate basis functions, but probably do not want all of them in the dictionary. Therefore, we seek a parsimonious model
  - to pick few entries from the dictionary and learn residual by deep learning.
- Appropriate treatment of noise . . . [Rudy/Kutz/Brunton '19]
- Missing/corrupted data in time series.
- Working with several applications in material science and chemical engineering.

Thank you for your attention!!

Selected References

Benner, P., Goyal, P., Kramer, B., Peherstorfer, B., and Willcox, K. (2020).
Operator inference for non-intrusive model reduction of systems with non-polynomial nonlinear terms.
Comp. Meth. Appl. Mech. Eng., 372:113433.

Brunton, S. L., Proctor, J. L., and Kutz, J. N. (2016).
Discovering governing equations from data by sparse identification of nonlinear dynamical systems.
Proc. Nat. Acad. Sci. U.S.A., 113(15):3932–3937.

Goyal, P. and Benner, P. (2021).
LQResNet: A deep neural network architecture for learning dynamic processes.
arXiv preprint arXiv:2103.02249.

He, K., Zhang, X., Ren, S., and Sun, J. (2016).
Deep residual learning for image recognition.
In *Proc. IEEE Conf. Comp. Vision Patt. Recog.*, pages 770–778.

Li, H., Xu, Z., Taylor, G., Studer, C., and Goldstein, T. (2018).
Visualizing the loss landscape of neural nets.
In *Neural Inform. Processing Syst.*

Peherstorfer, B. and Willcox, K. (2016).
Data-driven operator inference for nonintrusive projection-based model reduction.
Comp. Meth. Appl. Mech. Eng., 306:196–215.

Qian, E., Kramer, B., Peherstorfer, B., and Willcox, K. (2020).
Lift & learn: Physics-informed machine learning for large-scale nonlinear dynamical systems.
Physica D: Nonlinear Phenomena, 406:132401.

Rudy, S. H., Kutz, J. N., and Brunton, S. L. (2019).
Deep learning of dynamics and signal-noise decomposition with time-stepping constraints.
J. Comput. Phys., 396:483–506.

Out now — a Trilogy on Model Order Reduction