Lift & Learn: Physics-informed machine learning for large-scale nonlinear dynamical systems

Elizabeth Qian (elizqian@mit.edu)
Department of Aeronautics & Astronautics
Massachusetts Institute of Technology

Boris Kramer
Department of Mechanical & Aerospace Engineering
University of California - San Diego

Benjamin Peherstorfer
Courant Institute of Mathematical Sciences
New York University

Karen Willcox
Oden Institute for Computational Engineering & Sciences
University of Texas at Austin

Deriving low-dimensional models

Traditional solvers for nonlinear PDEs are expensive: need inexpensive surrogate models for practical computations

- Projection-based reduced models rely on full knowledge of physics and their construction traditionally requires intrusive access to codes
- Data-fit models in machine learning treat solvers as black boxes and ignore physics

We propose **Lift & Learn**, a physics-informed method for learning reduced models that can recover the generalization accuracy of traditional intrusive reduced models

- Knowledge of the governing PDE is exploited to identify a lifting map (variable transformation + auxiliary variables) that exposes quadratic structure in the PDE
- Lifting lets us reformulate nonlinear model reduction as a non-intrusive polynomial operator inference

Lifting PDEs to quadratic form

Consider the general nonlinear governing PDE with state s:

$$\frac{\partial s}{\partial t} = f(s)$$

A quadratic lifting map $\mathcal{T}$ transforms and augments the system state so that the PDE in the lifted state, $w = \mathcal{T}(s)$, contains only quadratic nonlinearities, e.g.:

$$\frac{\partial w}{\partial t} = a_0 w + a_1 \frac{\partial w}{\partial x} + a_2 w^2 + a_3 w \frac{\partial w}{\partial x}$$

This structure allows us to reformulate the learning task as a polynomial operator inference problem.

Example

Original PDE	*Lifting map*	*Lifted PDE*
$\frac{\partial s}{\partial t} = -e^s$	$\mathcal{T}: s \mapsto \begin{pmatrix} s \\ -e^s \end{pmatrix} \equiv \begin{pmatrix} w_1 \\ w_2 \end{pmatrix}$	$\frac{\partial}{\partial t}\begin{pmatrix} w_1 \\ w_2 \end{pmatrix} = \begin{pmatrix} w_2 \\ (w_2)^2 \end{pmatrix}$

How general is the lifting approach?

Many nonlinear terms in engineering applications can be lifted to quadratic form.[1] In some cases, quadratic transformations are known, e.g., the specific volume variables for the Euler and Navier-Stokes equations underlying many fluids applications.

How is the lifting derived?

Our current strategy is problem-specific: we introduce auxiliary variables for non-quadratic terms of the PDE and augment the system with evolution equations for these new variables. Automated discovery of a lifting is a direction for future work.

Acknowledgments

The authors acknowledge the NSF GRFP, the Hertz Foundation, the US Air Force (FA9550-17-1-0195, FA9550-15-1-0038, FA9550-18-1-0023), the US Department of Energy Applied Mathematics Program (DE-FG02-08ER2585, DE-SC0009297, DESC0019334), and the SUTD-MIT International Design Centre.

Lift & Learn

1. Solve N-dimensional spatial discretization of original nonlinear PDE to generate K snapshots. Apply **lifting** map to snapshot data to obtain K lifted state and time derivative pairs $(\mathbf{W}, \dot{\mathbf{W}} \in \mathbb{R}^{N' \times K})$
2. Compute a d-dimensional global basis, $\mathbf{V}_d$, for the lifted data, e.g. via Proper Orthogonal Decomposition (POD) and project data:
$$\widehat{\mathbf{W}} = \mathbf{V}_d^\top \mathbf{W}, \qquad \dot{\widehat{\mathbf{W}}} = \mathbf{V}_d^\top \dot{\mathbf{W}}$$
where $d \ll N$. The reduced model for projected data can be parameterized by small, dense matrix operators:
$$\frac{\mathrm{d}\hat{\mathbf{w}}}{\mathrm{d}t} = \hat{\mathbf{A}}\hat{\mathbf{w}} + \hat{\mathbf{H}}(\hat{\mathbf{w}} \otimes \hat{\mathbf{w}})$$
3. Use least-squares operator inference[2] procedure to **learn** $\hat{\mathbf{A}} \in \mathbb{R}^{d\times d}$, $\hat{\mathbf{H}} \in \mathbb{R}^{d \times d^2}$ from data:
$$\min_{\hat{\mathbf{A}}\in\mathbb{R}^{d\times d},\hat{\mathbf{H}}\in\mathbb{R}^{d\times d^2}} \frac{1}{K}\left\| \widehat{\mathbf{W}}^\top \hat{\mathbf{A}}^\top + \left(\widehat{\mathbf{W}} \otimes \widehat{\mathbf{W}}\right)^\top \hat{\mathbf{H}}^\top - \dot{\widehat{\mathbf{W}}}^\top \right\|_F^2$$

Bounding the residual of Lift & Learn models[3]:

If the original nonlinear PDE solver uses a spatial discretization scheme with order of accuracy p, and the map $\mathcal{T}$ is continuous with Lipschitz derivative, then the residual of the Lift & Learn model on the training data is bounded:

$$\min_{\hat{\mathbf{A}}\in\mathbb{R}^{d\times d},\hat{\mathbf{H}}\in\mathbb{R}^{d\times d^2}} \frac{1}{K}\left\| \widehat{\mathbf{W}}^\top \hat{\mathbf{A}}^\top + \left(\widehat{\mathbf{W}} \otimes \widehat{\mathbf{W}}\right)^\top \hat{\mathbf{H}}^\top - \dot{\widehat{\mathbf{W}}}^\top \right\|_F^2 \leq (c_0 N'^{0.5-p} + c_1 \varepsilon)^2$$

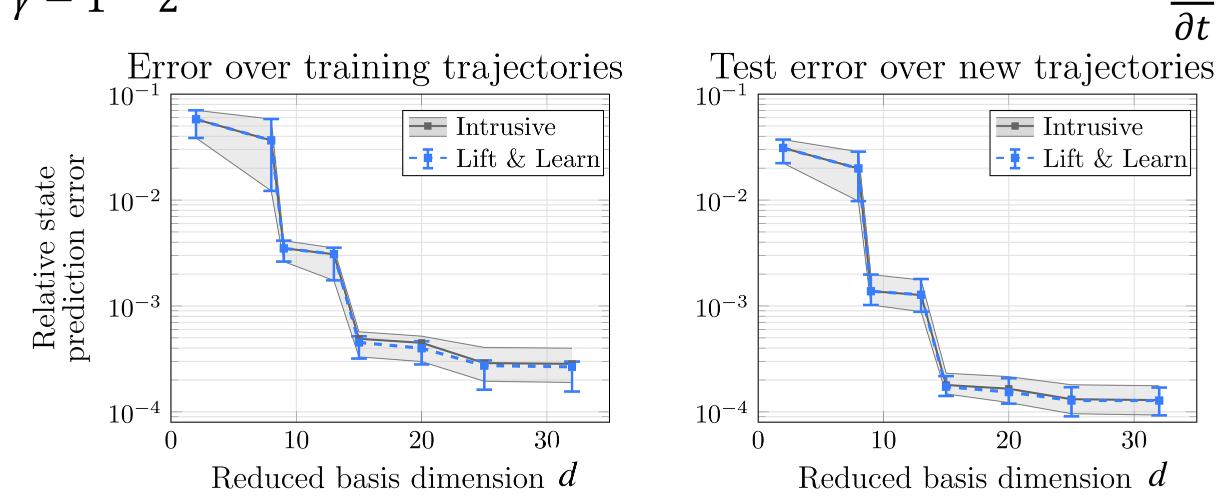
where ε is the projection error of $\mathbf{W}$ onto $\mathbf{V}_d$ and c_0, c_1 are constants.

Implications:

1. By respecting the problem physics in the lifted coordinates we can put an upper bound on the residual of the Lift & Learn model.
2. The Lift & Learn model residual is at least as good as the residual of an intrusive lifted POD reduced model.

Generalization & accuracy: Euler equations

Original PDE	*Lifting map*	*Lifted PDE*
$\frac{\partial}{\partial t}\begin{pmatrix} \rho \\ \rho u \\ E \end{pmatrix} + \frac{\partial}{\partial x}\begin{pmatrix} \rho u \\ \rho u^2 + p \\ (E+p)u \end{pmatrix} = 0$, $E = \frac{p}{\gamma - 1} + \frac{1}{2}\rho u^2$	$\mathcal{T}: \begin{pmatrix} \rho \\ \rho u \\ E \end{pmatrix} \mapsto \begin{pmatrix} 1/\rho \equiv \zeta \\ u \\ p \end{pmatrix}$	$\frac{\partial \zeta}{\partial t} = -u\frac{\partial \zeta}{\partial x} + \zeta\frac{\partial u}{\partial x}$, $\frac{\partial u}{\partial t} = -u\frac{\partial u}{\partial x} - \zeta\frac{\partial p}{\partial x}$, $\frac{\partial p}{\partial t} = -\gamma p\frac{\partial u}{\partial x} - u\frac{\partial p}{\partial x}$



Generalization & accuracy: FitzHugh-Nagumo neuron activation model[4]

A benchmark problem in nonlinear model reduction:

Original PDE:

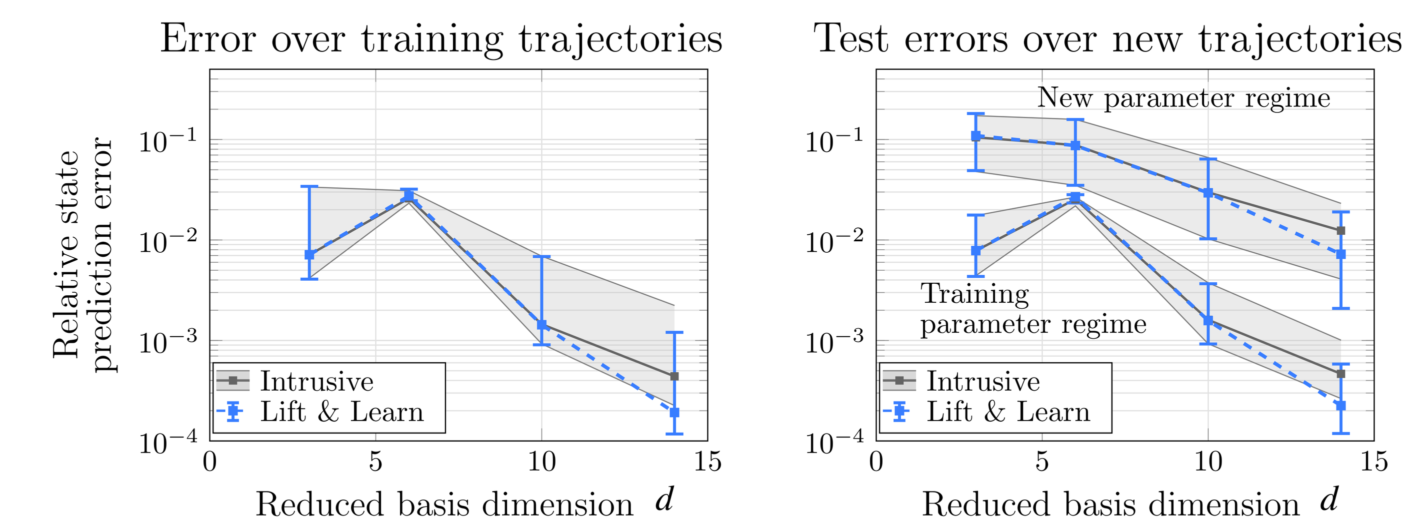
$$\gamma\frac{\partial s_1}{\partial t} = \gamma^2\frac{\partial^2 s_1}{\partial x^2} - s_1^3 + 1.1s_1^2 - 0.1s_1 + s_2 + 0.05$$
$$\frac{\partial s_2}{\partial t} = 0.5s_1 - 2s_2 + 0.05$$

Lifting map:

$$\mathcal{T}: \begin{pmatrix} s_1 \\ s_2 \end{pmatrix} \mapsto \begin{pmatrix} s_1 \\ s_2 \\ (s_1)^2 \end{pmatrix} \equiv \begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix}$$

Lifted quadratic PDE:

$$\gamma\frac{\partial w_1}{\partial t} = \gamma^2\frac{\partial^2 w_1}{\partial x^2} - w_1 w_3 + 1.1w_1^2 - 0.1w_1 + w_2 + 0.05$$
$$\frac{\partial w_2}{\partial t} = 0.5w_1 - 2w_2 + 0.05$$
$$\frac{\gamma}{2}\frac{\partial w_3}{\partial t} = \gamma^2 w_1 \frac{\partial^2 w_1}{\partial x^2} - w_3^2 + 1.1w_1 w_3 - 0.1w_3 + w_1 w_2 + 0.05w_1$$



Conclusions

Our Lift & Learn approach infers low-dimensional quadratic models for nonlinear PDEs:

- Lifting maps expose quadratic structure in the nonlinear PDE so that a low-dimensional model can be explicitly parametrized by polynomial matrix operators
- We fit polynomial operators to lifted data obtained non-intrusively from the original nonlinear model
- Numerical experiments show that Lift & Learn models recover the generalization accuracy of intrusive projection-based reduced models

References

[1] C. Gu, QLMOR: A projection-based nonlinear model order reduction approach using quadratic-linear representation of nonlinear systems, IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst. 30 (2011) 1307–1320.
[2] B. Peherstorfer and K. Willcox, Data-driven operator inference for nonintrusive projection-based model reduction, Comput. Method Appl. M., 306 (2016), pp. 196–215.
[3] E. Qian, B. Kramer, B. Peherstorfer, and K. Willcox, "Lift & Learn: Physics-informed machine learning for large-scale nonlinear dynamical systems." Physica D: Nonlinear Phenomena 406 (2020), pp. 132401
[4] S. Chaturantabut, D. Sorensen, Nonlinear model reduction via discrete empirical interpolation, SIAM J. on Sci. Comp. 32 (2010) 2737–2764.