Reduced Order Modeling using Shallow ReLU Networks with Grassmann Layers

Kayla Bollinger and Hayden Schaeffer
Carnegie Mellon University

August 16-19th, 2021
MSML21: Mathematical and Scientific Machine Learning

Problem Statement

Approximate:

$$f : \Omega \subseteq \mathbb{R}^m \to \mathbb{R}^n \text{ where } m \gg 1 \text{ , given } \{(x_\ell, f(x_\ell), Df(x_\ell))\}_{\ell=1}^{M}$$

Application Examples

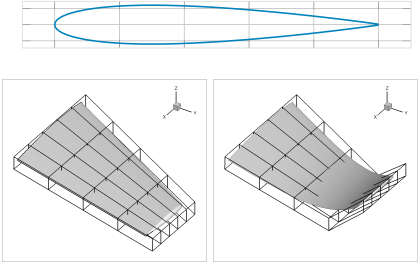


Figure 1: Top: **NACA0012 airfoil**, $m = 18$ (Hicks-Henne shape parameters); Bottom: **ONERA-M6 wing**, $m = 50$ (free-form deformation parameters) (image source: Lukaczyk et al. 2014); $f =$ lift/drag coefficients

Table of contents

Overview

Reduced Order Modeling (ROM)

ROMs: approximate high-dimensional complex systems by simpler low-dimensional systems

Examples:

- proper orthogonal decomposition (POD), e.g. Berkooz, Holmes, and Lumley 1993; Holmes et al. 2012
- global sensitivity analysis, e.g. Saltelli et al. 2008
- sliced inverse regression, e.g. Li 1991
- compressed sensing, e.g. Fornasier, Schnass, and Vybiral 2012
- *active subspace*, e.g. Russi 2010; Constantine, Dow, and Wang 2014

Illustrative Function for Active Subspace

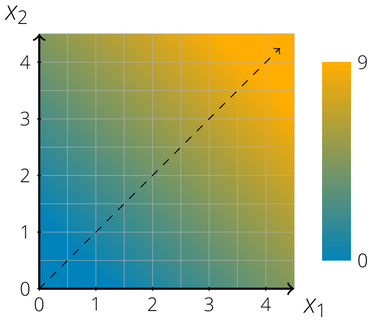


Figure 2: $f(x_1, x_2) = x_1 + x_2$

Active Subspace

Consider:

$$f \in C^1(\Omega; \mathbb{R}^n), \quad \Omega \subseteq \mathbb{R}^m \quad ; \quad C \in \mathbb{R}^{m\times m} \text{ defined by } C := \mathbb{E}\left[Df^T Df\right] = W\Lambda W^T$$

$$W = \Big[\overbrace{W_1}^{m\times k} \quad \underbrace{W_2}_{m\times(m-k)}\Big], \quad \Lambda = \begin{bmatrix} \overbrace{\Lambda_1}^{k\times k} & \\ & \underbrace{\Lambda_2}_{(m-k)\times(m-k)} \end{bmatrix}, \quad (\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m)$$

Lemma (based on Lemma 2.2[1])

Under the rotated coordinates $y = W_1^T x \in \mathbb{R}^k$ and $z = W_2^T x \in \mathbb{R}^{m-k}$:

$$\sum_{i=1}^{n} \mathbb{E}[\nabla_y(f_i)^T \nabla_y(f_i)] = \lambda_1 + \cdots + \lambda_k$$

$$\sum_{i=1}^{n} \mathbb{E}[\nabla_z(f_i)^T \nabla_z(f_i)] = \lambda_{k+1} + \cdots + \lambda_m$$

[1] Constantine, Dow, and Wang 2014

Methodology

Surrogate Model

Given:

$$f : \Omega \subseteq \mathbb{R}^m \to \mathbb{R}^n, \text{ with } m \gg 1$$

Surrogate Model:

$$f \approx g_\theta \circ U^T$$

where

- $U \in \mathbb{R}^{m \times k}$; maps input space to k-dimensional subspace
  - construction inspired by the *active subspace* method, i.e. utilizing the dominate left singular vectors of the matrix

$$\left[Df(x_1)^T \ldots Df(x_M)^T \right]$$

- $g_\theta : \mathbb{R}^k \to \mathbb{R}^n$; approximates f with respect to k- dimensional inputs
  - represented by a *shallow ReLU network* with hidden dimension h and trainable parameters $\theta \in \mathbb{R}^d$ ($d = h(k + n + 1) + n$)

$$g_{\theta(y)} = A_2(\text{ReLU}(A_1 y + b_1)) + b_2$$

Alternating Minimization

$$\min_{\theta \in \mathbb{R}^d,\ \mathrm{Range}\, U \in Gr(k,m)} \frac{1}{M} \sum_{\ell=1}^{M} \| f(x_\ell) - g_\theta(U^T x_\ell) \|_2^2 + \lambda \|\theta\|_2^2.$$

Algorithm 1 (alternating minimization scheme)

Initialize: U via active subspace method, θ for ReLU network g, learning rate τ.

repeat

(1) Update weights θ via ADAM method applied to:

$$\min_{\theta \in \mathbb{R}^d} \frac{1}{M} \sum_{\ell=1}^{M} \| f(x_\ell) - g_\theta(U^T x_\ell) \|_2^2 + \lambda \|\theta\|_2^2$$

(2) Update U via the Grassmann manifold-constrained least squares problem:

$$\min_{\mathrm{Range}\, U \in Gr(k,m)} \frac{1}{M} \sum_{\ell=1}^{M} \| f(x_\ell) - g_\theta(U^T x_\ell) \|_2^2$$

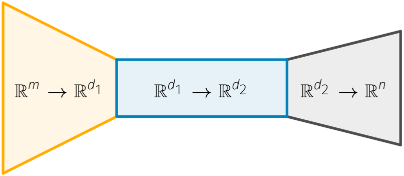
(3) Update: $\tau = 0.9\tau$

until converged

Other Neural Network (NN) based ROMs

- POD method with NN applied to fluid flows, Hesthaven and Ubbiali 2018; Lui and Wolf 2019
- PCA-based approach with NN applied to PDEs, Bhattacharya et al. 2020
- NN used to pick models from a dictionary of local ROMs, Daniel et al. 2020
- manifold learning using deep NN, Shaham, Cloninger, and Coifman 2018; Chui and Mhaskar 2018; Zhu et al. 2018
- encoder/decoder structured NN, Hinton and Salakhutdinov 2006; Ravi 2017
- *active subspace for input reduction and POD for output reduction with NN nonlinear fit in-between*, O'Leary-Roseberry et al. 2020

Example Encoder/Decoder Structure



Experimental Results

Strength in Data Scarce Setting

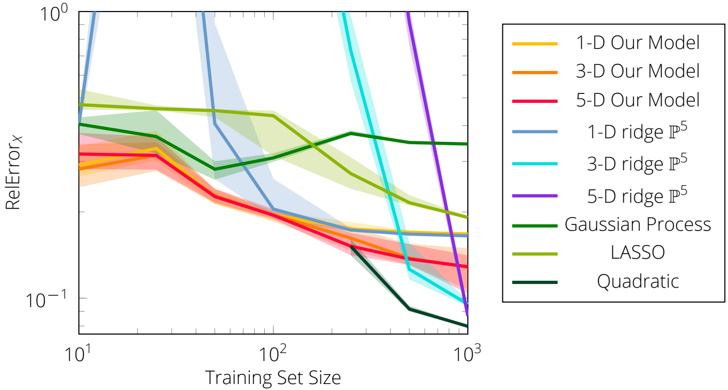
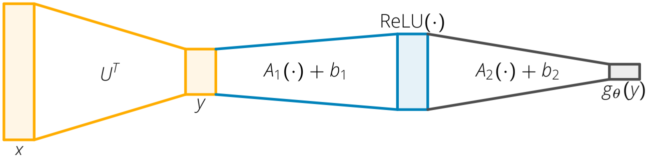


Figure 3: NACA0012 drag coefficient data; $\text{RelError}_X = \left(\frac{\sum_{x \in X} \|f(x) - \tilde{f}(x)\|_2^2}{\sum_{x \in X} \|f(x)\|_2^2} \right)^{\frac{1}{2}}$, $\tilde{f}$ represents the trained model for a given approach.

Benefits of Additional Structure/Interpretability



Our Model

	h=8	h=64	h=256
k=1	0.41	0.41	0.48
k=2	0.41	0.37	0.37
k=3	0.40	0.37	0.33

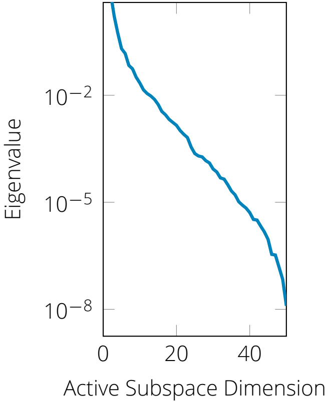
Bowtie Model

	h=8	h=64	h=256
k=1	0.47	0.47	0.47
k=2	0.50	0.48	0.52
k=3	0.49	0.51	0.51

Table 1: Relative validation error $\left(\frac{\frac{1}{|X_{Val}|} \sum_{x \in X_{Val}} \|f(x) - \tilde{f}(x)\|_2^2}{\frac{1}{|X|} \sum_{x \in X} \|f(x)\|_2^2} \right)^{\frac{1}{2}}$ of models applied to NACA0012 drag coefficient data; training/validation set size of 50.

Application to ONERA-M6 Wing

Eigenvalues (Λ) Show Potential
for Accuracy in Low Dimensions



	h=8	h=64	h=256
k=1	0.12	0.12	0.15
k=2	0.11	0.09	0.10
k=3	0.11	0.08	0.08

Table 2: Relative validation error of *our model* applied to **ONERA-M6 drag** coefficient data; training/validation set size of 50.

Acknowledgements

Support

This research was partly funded by:

- AFOSR, FA9550-17-1-0125
- NSF CAREER grant #1752116

References

Berkooz, Gal, Philip Holmes, and John L Lumley (1993). “The proper orthogonal decomposition in the analysis of turbulent flows”. In: *Annual review of fluid mechanics* 25.1, pp. 539–575.

Bhattacharya, Kaushik et al. (2020). “Model reduction and neural networks for parametric pdes”. In: *arXiv preprint arXiv:2005.03180*.

Chui, Charles K. and Hrushikesh N. Mhaskar (2018). “Deep Nets for Local Manifold Learning”. In: *Frontiers in Applied Mathematics and Statistics* 4, p. 12. ISSN: 2297-4687. DOI: `10.3389/fams.2018.00012`. URL: `https://www.frontiersin.org/article/10.3389/fams.2018.00012`.

Constantine, Paul G., Eric Dow, and Qiqi Wang (Jan. 2014). “Active Subspace Methods in Theory and Practice: Applications to Kriging Surfaces”. In: *SIAM Journal on Scientific Computing* 36.4. ISSN: 1095-7197. DOI: `10.1137/130916138`. URL: `http://dx.doi.org/10.1137/130916138`.

Daniel, Thomas et al. (Dec. 2020). “Model order reduction assisted by deep neural networks (ROM-net)”. In: *Advanced Modeling and Simulation in Engineering Sciences* 7. DOI: `10.1186/s40323-020-00153-6`.

Fornasier, Massimo, Karin Schnass, and Jan Vybiral (2012). “Learning functions of few arbitrary linear parameters in high dimensions”. In: *Foundations of Computational Mathematics* 12.2, pp. 229–262.

Hesthaven, J.S. and S. Ubbiali (2018). “Non-intrusive reduced order modeling of nonlinear problems using neural networks”. In: *Journal of Computational Physics* 363, pp. 55–78. ISSN: 0021-9991. DOI: https://doi.org/10.1016/j.jcp.2018.02.037. URL: https://www.sciencedirect.com/science/article/pii/S0021999118301190.

Hinton, G. E. and R. R. Salakhutdinov (2006). “Reducing the Dimensionality of Data with Neural Networks”. In: *Science* 313.5786, pp. 504–507. ISSN: 0036-8075. DOI: 10.1126/science.1127647. eprint: https://science.sciencemag.org/content/313/5786/504.full.pdf. URL: https://science.sciencemag.org/content/313/5786/504.

Holmes, Philip et al. (2012). *Turbulence, coherent structures, dynamical systems and symmetry*. Cambridge university press.

Li, Ker-Chau (1991). “Sliced inverse regression for dimension reduction”. In: *Journal of the American Statistical Association* 86.414, pp. 316–327.

Lui, Hugo F. S. and William R. Wolf (June 2019). “Construction of reduced-order models for fluid flows using deep feedforward neural networks”. In: *Journal of Fluid Mechanics* 872, pp. 963–994. ISSN: 1469-7645. DOI: `10.1017/jfm.2019.358`. URL: `http://dx.doi.org/10.1017/jfm.2019.358`.

Lukaczyk, Trent W. et al. (2014). “Active Subspaces for Shape Optimization”. In: *10th AIAA Multidisciplinary Design Optimization Conference*. DOI: `10.2514/6.2014-1171`. eprint: `https://arc.aiaa.org/doi/pdf/10.2514/6.2014-1171`. URL: `https://arc.aiaa.org/doi/abs/10.2514/6.2014-1171`.

O’Leary-Roseberry, Thomas et al. (2020). *Derivative-Informed Projected Neural Networks for High-Dimensional Parametric Maps Governed by PDEs*. arXiv: `2011.15110 [math.NA]`.

Ravi, Sujith (2017). *ProjectionNet: Learning Efficient On-Device Deep Networks Using Neural Projections*. arXiv: `1708.00630 [cs.LG]`.

Russi, Trent Michael (2010). “Uncertainty Quantification with Experimental Data and Complex System Models”. PhD thesis. University of California, Berkeley.

Saltelli, Andrea et al. (2008). *Global sensitivity analysis: the primer*. John Wiley & Sons.

Shaham, Uri, Alexander Cloninger, and Ronald R. Coifman (2018). “Provable approximation properties for deep neural networks”. In: *Applied and Computational Harmonic Analysis* 44.3, pp. 537–557. ISSN: 1063-5203. DOI: https://doi.org/10.1016/j.acha.2016.04.003. URL: https://www.sciencedirect.com/science/article/pii/S1063520316300033.

Zhu, W. et al. (June 2018). “LDMNet: Low Dimensional Manifold Regularized Neural Networks”. In: *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2743–2751. DOI: 10.1109/CVPR.2018.00290.

(Virtual) Questions?

Email: kbolling@andrew.cmu.edu

Code: https://github.com/kaylabollinger/ROM_AS-NN.git