Clues for noise robustness of state estimation: Error-curve quest of neural network and linear regression

Taichi Nakamura[1], Kai Fukami[2,1], Koji Fukagata[1]

[1] Mechanical Engineering, Keio University (Japan), [2] Mechanical and Aerospace Engineering, University of California, Los Angeles

Motivation

Neural networks in science & engineering

[1]

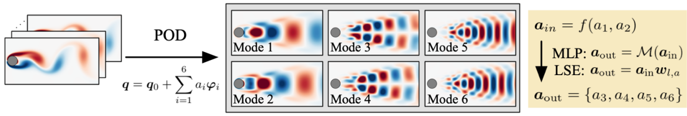
- NNs have acquired citizenship as a surrogate for linear methods in science and engineering thanks to capabilities to consider nonlinearity
- Limitation of NNs: lack of interpretability
  ⇒ Linear methods are still indispensable for their transparency
- Our interest: Can we obtain any hints from interpretable linear methods to improve the practicability of NNs?

[1] https://t3.ftcdn.net/jpg/01/90/18/04/360_F_190180455_Jmyl2Mq VogM80zFehWZ49rAM65il8AaH.jpg

MLP vs linear stochastic estimation

- Consider a canonical fluid flow regression problem to compare multi-layer perceptron (MLP) and linear stochastic estimation (LSE)
- Also focus on noise robustness and perform error-curve analysis

Problem setting



- Take proper orthogonal decomposition (POD) for a flow around two-dimensional cylinder at $Re_D = 100$ [2]

[2] Kor et al., *J Fluid Sci. Technol.*, 2017

- Training configuration
  - Input: low-order POD coefficients $\boldsymbol{a}_{\rm in} = f(a_1, a_2)$
  - Output: high-order POD coefficients $\boldsymbol{a}_{\rm out} = \{a_3, a_4, a_5, a_6\}$
- The POD coefficients a_i can be approximated in Fourier forms
- High-order POD coefficients can be represented by the quadratic expression of a_1 and a_2 due to its triadic interaction [3][4]
- Estimation models

[3] Loiseau et al., *Applications*, 2020
[4] Nair et al., *Phys. Rev. E*, 2018

  - MLP: aggregate of a minimum unit called `perceptron'

$$\boldsymbol{w}_{\rm MLP} = {\rm argmin}_{\boldsymbol{w}_{\rm MLP}} ||\boldsymbol{a}_{\rm out} - \mathcal{M}(\boldsymbol{a}_{\rm in}; \boldsymbol{w}_{\rm MLP})||_2$$

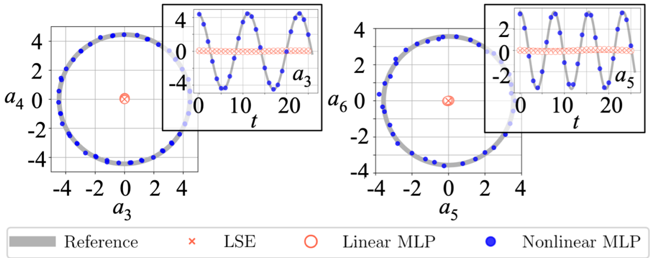
  - LSE: express output data as a linear map of input data

$$\boldsymbol{w}_{\rm LSE} = {\rm argmin}_{\boldsymbol{w}_{\rm LSE}} ||\boldsymbol{a}_{\rm out} - \boldsymbol{a}_{\rm in}\boldsymbol{w}_{\rm LSE}||_2 = ((\boldsymbol{a}_{\rm in}\boldsymbol{w}_{\rm LSE})^T \boldsymbol{a}_{\rm in}\boldsymbol{w}_{\rm LSE})^{-1}(\boldsymbol{a}_{\rm in}\boldsymbol{w}_{\rm LSE})^T \boldsymbol{a}_{\rm out}$$

Results

POD coefficients estimation

- 1st order input $\boldsymbol{a}_{\rm in} = \boldsymbol{a}^{\rm 1st} = \{a_1, a_2\}$
  - LSE vs linear MLP vs nonlinear MLP



[L_2 error norm] LSE: 1.00, linear MLP: 1.00, nonlinear MLP: 0.0119

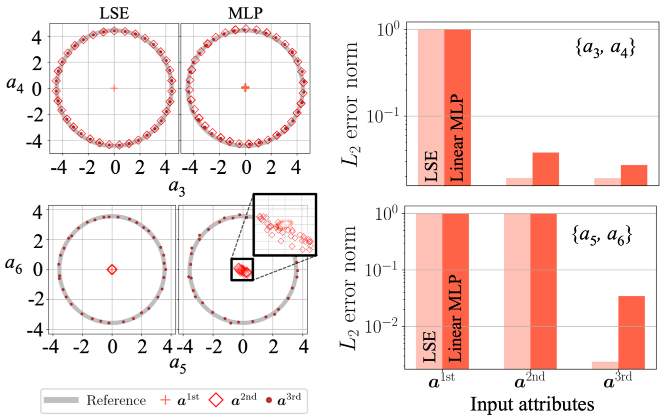
  - Nonlinear activation function works well for estimation

- 2nd/3rd order input $\boldsymbol{a}_{\rm in} = \boldsymbol{a}^{\rm 2nd}$ or $\boldsymbol{a}^{\rm 3rd}$

$$\boldsymbol{a}^{\rm 2nd} = \{a_1, a_2, a_1a_2, a_1^2, a_2^2\}$$

$$\boldsymbol{a}^{\rm 3rd} = \{a_1, a_2, a_1a_2, a_1^2, a_2^2, a_1^2a_2, a_1a_2^2, a_1^3, a_2^3\}$$

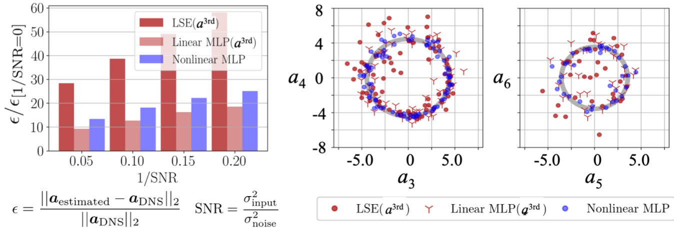
  - LSE vs linear MLP



  - The output $\{a_3, a_4\}$ requires the input up to 2nd order term $a^{\rm 2nd}$
  - The output $\{a_5, a_6\}$ requires the input up to 3rd order term $a^{\rm 3rd}$
  - Nonlinearity can be replaced by giving a proper data input

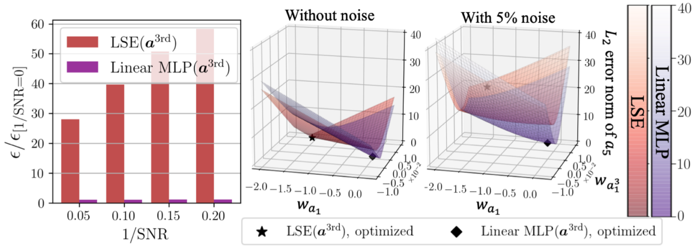
Noise robustness

- Perturb the input with Gaussian noise
  - Input: LSE, linear MLP → $\boldsymbol{a}^{\rm 2nd}$ nonlinear MLP → $\boldsymbol{a}^{\rm 3rd}$



  - The response of the LSE is more sensitive than that of the MLPs
  - *What does contribute to noise robustness in the linear MLP?*

- Factor contributing to the noise robustness
  - Shallow linear MLP model: no middle layer, no bias
  - Increase ratio of the L_2 error norm & error surfaces of output



  - Differences in optimized values are caused by optimization methods choice (LSE: least squares method, MLP: gradient method)
  - The noise drastically deforms the error surface of the LSE

Conclusions

- The differences between MLP and linear stochastic estimation (LSE) was investigated by considering a fluid flow regression problem
- Efficacy of nonlinear activation can be observed
- Noise robustness with error-curve analysis
  - The linear MLP was more robust for noise than the LSE
  - Optimization method contributed to the noise robustness
  - Noise robustness was visualized by using error surface