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 Department of Mechanical Engineering, Keio University
2 Department of Mechanical and Aerospace Engineering, University of California, Los Angeles

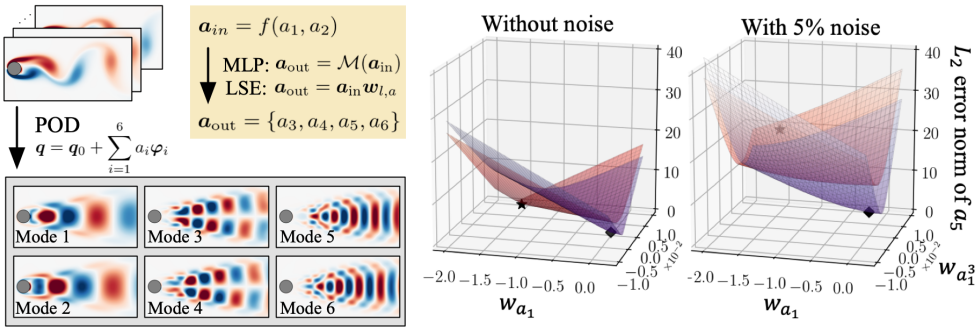

Figure 1: Error-curve analysis for POD coefficient estimation.

ABSTRACT

Neural network (NN), which is known as a universal approximator, has been utilized for a wide range of engineering purposes such as state estimation, anomaly detection, and system control. However, it is also the fact that conventional linear methods are still indispensable for them because of their uniqueness and interpretability of modeling results. Of particular interest here is whether we can gain some clues to develop more advanced NN techniques by comparing them to the linear models in a fundamental manner. We here consider a canonical fluid flow regression problem based on a proper orthogonal decomposition for an unsteady fluid flow. Concretely, we compare a linear stochastic estimation and a multi-layer perceptron from the perspective of the response on the error curve constructed by weights inside the models when we add noisy perturbation to the input data. Our analysis clearly visualizes the "robustness" against noise on the error-curve domain.

1 INTRODUCTION

Linear-theory-based analyses have contributed to the development of science and engineering to date (Brunton & Kutz, 2019). One of the beauties of them is the generalizability and the transparency of results provided by the models, which enables us toward practical application. However, there is still a big hindrance to their uses since the real nature is governed by strong nonlinearities in both space and time. Nonlinear neural networks (NNs) have recently acquired citizenship as one of the considerable surrogates, although a black-box use of NNs is treated as the conundrum for practical applications. Our idea here is to borrow some hints from the interpretable linear methods to improve the practicability of NNs.

In this paper, we consider a canonical fluid flow regression problem to compare a linear stochastic estimation and a multi-layer perceptron from the perspective on noise robustness. The present models attempt to estimate high-order proper orthogonal decomposition (POD) coefficients from low-order counterparts of a flow around a two-dimensional cylinder (Loiseau et al., 2018), which is a simple problem but contains a strong nonlinear input-output relationship. In particular, our analyses are founded on the observation on the error surfaces constructed by weights inside the models.

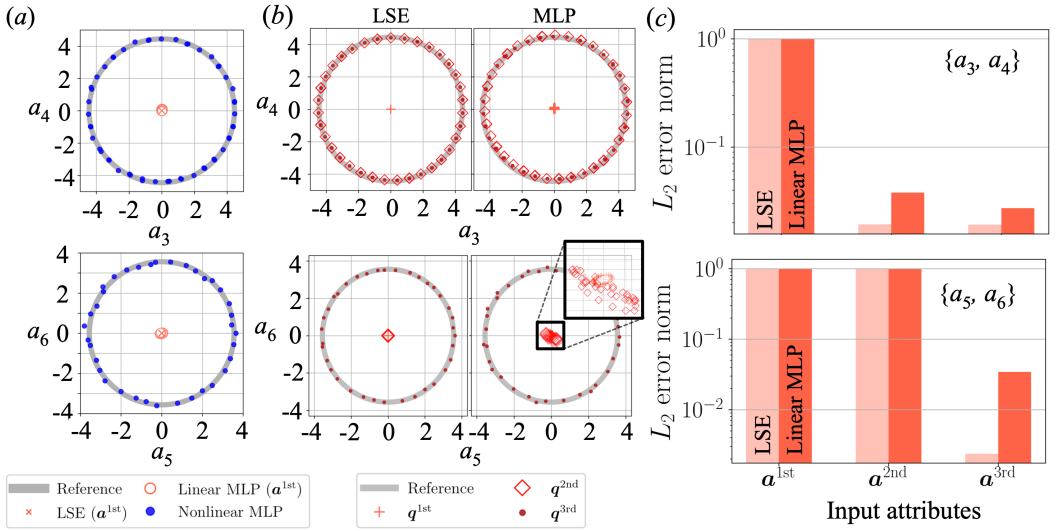

Figure 2: POD coefficients estimation from (a) 1st order coefficients $\boldsymbol{a}_{\text{in}} = \{a_1, a_2\}$ and (b) 2nd & 3rd order coefficients, and (c) the L_2 error norm.

2 METHODS

2.1 PROBLEM SETTING

We consider the estimation of high-order POD coefficients $\boldsymbol{a}_{\text{out}} = \{a_3, a_4, a_5, a_6\}$ of flow around two-dimensional cylinder at $Re_D = 100$ from the information of low-order counterparts $\boldsymbol{a}_{\text{in}} = f(a_1, a_2)$ (Loiseau et al., 2018). Although the problem setting is quite low dimension thanks to the assistance by POD, let us emphasize that this problem setting holds a strong nonlinearity because the relationship between $\boldsymbol{a}_{\text{in}}$ and $\boldsymbol{a}_{\text{out}}$ is governed by the triadic interactions arising from the nonlinear term of the Navier–Stokes equations. Due to this relationship, Loiseau et al. (2018) reported that high-order POD coefficients $\boldsymbol{a}_{\text{out}}$ can be estimated with a linear superposition of high-order terms of low-order POD coefficients, e.g., a_1^2, $a_1 a_2$, and $a_1^2 a_2^2$. Using this setup, we aim to reveal the fundamental difference of linearity and nonlinearity inside models. The problem can mathematically be formed as $\boldsymbol{a}_{\text{out}} = \mathcal{F}(\boldsymbol{a}_{\text{in}})$, where $\mathcal{F}$ denotes an estimation model. As the model $\mathcal{F}$, we use the LSE and the MLP, which will be described later. The flow snapshots are prepared with a two-dimensional direct numerical simulation by numerically solving incompressible continuity and Navier-Stokes equations (Kor et al., 2017). The POD is then taken for the collected snapshots to decompose the flow field $\boldsymbol{q}$ as $\boldsymbol{q} = \boldsymbol{q}_0 + \sum_{i=1}^{M} a_i \boldsymbol{\varphi}_i$, where $\boldsymbol{\varphi}$ represents a POD basis, a is the POD coefficient, $\boldsymbol{q}_0$ is the temporal average of the flow field, and M denotes the number of POD modes. For training the MLP and LSE, we use 5000 snapshots.

2.2 MULTI-LAYER PERCEPTRON

We use a multi-layer perceptron (MLP) (Rumelhart et al., 1986) as an example of the neural networks. The present MLP contains hidden units of 4-8-16-8, while the number of output nodes is 4 (3rd to 6th POD modes). The number of input nodes varies depending on covered cases, which will be explained later. Since the present MLP $\mathcal{M}$ attempts to output $\boldsymbol{a}_{\text{out}} = \{a_3, a_4, a_5, a_6\}$ from the input $\boldsymbol{a}_{\text{in}} = f(a_1, a_2)$, the problem setting regarding weights $\boldsymbol{w}_m$ inside the MLP can be represented as

$$\boldsymbol{w}_m = \text{argmin}_{\boldsymbol{w}_m} ||\boldsymbol{a}_{\text{out}} - \mathcal{M}(\boldsymbol{a}_{\text{in}}; \boldsymbol{w}_m)||_2. \tag{1}$$

2.3 LINEAR STOCHASTIC ESTIMATION

As an example of the linear methods, we consider a linear stochastic estimation (LSE) (Adrian & Moin, 1988). The LSE expresses output data $\boldsymbol{Q} \in \mathbb{R}^{n_{\text{output}} \times n_{\text{data}}}$ as a linear map $\boldsymbol{w}_l \in \mathbb{R}^{n_{\text{input}} \times n_{\text{output}}}$ with respect to input data $\boldsymbol{P} \in \mathbb{R}^{n_{\text{input}} \times n_{\text{data}}}$ such that $\boldsymbol{Q} = \boldsymbol{w}_l \boldsymbol{P}$, where n_{data}

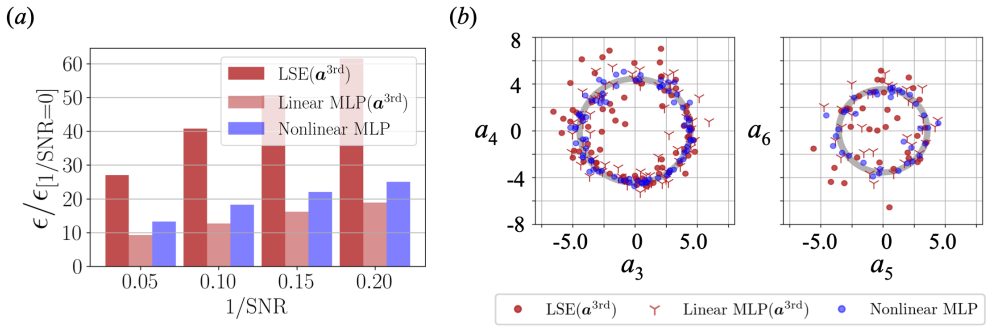

Figure 3: Robustness for noisy input. (a) Dependence of the increase ratio of the L_2 error norm $\epsilon/\epsilon_{[1/\mathrm{SNR}=0]}$ on noise magnitude. (b) $\boldsymbol{a}_{\mathrm{out}}$ with $1/\mathrm{SNR} = 0.1$ for the linear models with $\boldsymbol{a}_{\mathrm{in}} = \boldsymbol{a}^{3\mathrm{rd}}$ and the nonlinear MLP with $\boldsymbol{a}_{\mathrm{in}} = \boldsymbol{a}^{1\mathrm{st}}$.

is the number of training snapshots, n_{input} is the number of input attribute, and n_{output} is the number of output attribute. The linear map $\boldsymbol{w}_l$ is optimized through the minimization manner as

$$\boldsymbol{w}_l = \mathrm{argmin}_{\boldsymbol{w}_l} \parallel \boldsymbol{Q} - \boldsymbol{w}_l \boldsymbol{P} \parallel_2, \tag{2}$$

which is analogous to the optimization of NNs. Note that we do not include penalization terms. e.g., L_1 and L_2 penalties, in the present loss function for fair comparison to the covered NNs in terms of weight updates.

3 Results

We first compare the LSE to the MLP with linear activation and nonlinear ReLU function to examine whether nonlinearity works well for the estimation or not. Moreover, we consider three patterns of input $f(\boldsymbol{a}_1, \boldsymbol{a}_2)$ for the LSE and the linear MLP, while using only a_1 and a_2 with the nonlinear MLP. The input $f(\boldsymbol{a}_1, \boldsymbol{a}_2)$ is $\boldsymbol{a}^{1\mathrm{st}} = \{a_1, a_2\}$ or $\boldsymbol{a}^{2\mathrm{nd}} = \{a_1, a_2, a_1a_2, a_1^2, a_2^2\}$ or $\boldsymbol{a}^{3\mathrm{rd}} = \{a_1, a_2, a_1a_2, a_1^2, a_2^2, a_1^2a_2, a_1a_2^2, a_1^3, a_2^3\}$. This investigation enables us to check whether the linear models can also be utilized by giving a proper input or not, because Loiseau et al. (2018) reported that the high-order coefficients $\boldsymbol{a}_{\mathrm{out}}$ can be represented using the quadratic expression of a_1 and a_2, as mentioned above.

The estimated coefficients $\boldsymbol{a}_{\mathrm{out}} = \{a_3, a_4, a_5, a_6\}$ from only the information of the first-order coefficients $\boldsymbol{a}^{1\mathrm{st}} = \{a_1, a_2\}$ are shown in figure $2(a)$. Only the nonlinear MLP is in reasonable agreement with the reference in both coefficient maps, which reports the L_2 error norm $\epsilon = ||\boldsymbol{a}_{\mathrm{out,ref}} - \boldsymbol{a}_{\mathrm{out,est}}||_2/||\boldsymbol{a}_{\mathrm{out,ref}}||_2$ of 1.00 (LSE), 1.00 (linear MLP), and 0.0119 (nonlinear MLP). This suggests that a nonlinear activation function works effectively in estimation. However, this nonlinear influence can be replaced by using a proper input data set, i.e., $\boldsymbol{a}_{\mathrm{in}} = \{\boldsymbol{a}^{2\mathrm{nd}}, \boldsymbol{a}^{3\mathrm{rd}}\}$, even though we only use the linear methods, as shown in figures $2(b)$ and (c). By utilizing the input up to 2nd order terms $\boldsymbol{a}^{2\mathrm{nd}}$, the reasonable estimation for $\{a_3, a_4\}$ can be performed, while that for $\{a_5, a_6\}$ requires the input up to 3rd order terms $\boldsymbol{a}^{3\mathrm{rd}}$ with both the LSE and the linear MLP. This trend coincides with the fact that the POD coefficients of modes 3 and 4 can be expressed by the quadratic expression of a_1 and a_2, and that of modes 5 and 6 can be written as the cubic expression of a_1 and a_2 (Loiseau et al., 2018).

Although both linear methods employ well by giving a proper input as reported above, the LSE slightly outperforms the linear MLP for the high-order coefficient inputs — let us then compare these two methods in detail. To reveal the fundamental difference of the methods, we focus on the robustness against noisy input. We here introduce the white Gaussian noise defined by the signal-to-noise ratio (SNR), $\mathrm{SNR} = \sigma_{\mathrm{data}}^2/\sigma_{\mathrm{noise}}^2$, where σ_{data} and σ_{noise} respectively denote standard deviations of input data and noise. The responses to noisy inputs are summarized in figure 3. We use the LSE and the linear MLP with $\boldsymbol{a}_{\mathrm{in}} = \boldsymbol{a}^{3\mathrm{rd}}$ as the linear models, and the nonlinear MLP with $\boldsymbol{a}_{\mathrm{in}} = \boldsymbol{a}^{1\mathrm{st}}$ is also monitored for comparison. Notably, the response of LSE is much more sensitive than that of the covered MLPs, which shows the opposite behavior against the case without noise.

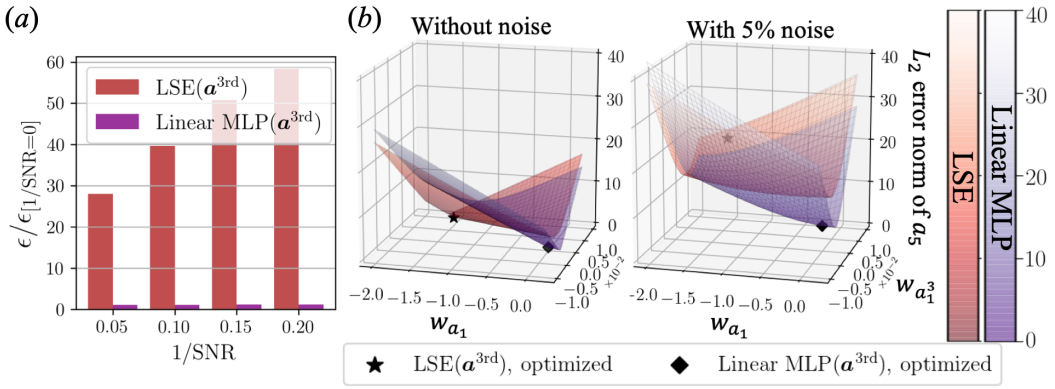

Figure 4: Robustness of linear methods for noisy input. (a) Dependence of the increase ratio of the L_2 error norm $\epsilon/\epsilon_{[1/\mathrm{SNR}=0]}$ on noise magnitude. (b) Error surfaces of output a_5.

To identify the factor which is responsible for noise robustness, we construct a new linear MLP model $\mathcal{M}'(\boldsymbol{a}_{\mathrm{in}}; \boldsymbol{w}_{m'})$, which has no middle layer and no bias, such that $\boldsymbol{w}_{m'} \in \mathbb{R}^{n_{\mathrm{input}} \times n_{\mathrm{output}}}$. This setting enables us to compare the MLP and the LSE fairly since the number of contained weights inside them is identical with each other. The response to noisy input here is shown in figure $4(a)$. Ever for this shallow settings, the MLP still shows advantage against the LSE in terms of noise robustness. To examine this point more, we visualize the error curve surface by picking two weights w_{a_1} and $w_{a_1^3}$ which respectively correspond to input a_1 and a_1^3 up, and visualizing the error surface of a_5, as shown in figure $4(b)$. As shown, the optimized solutions through the minimization manner of MLP and LSE are different with each other, which is likely due to the difference in optimization methods. Moreover, what is striking here is that the noise addition drastically changes the shape of the error surface of LSE, while that of MLP is just pushed up in the normal direction. It implies that the weights obtained by the LSE can guarantee the minimum loss over the training data; however, may not be the optimal solution from the viewpoint of noise robustness. In contrast, the location of minimum point with the MLP does not change too much against the case without noise, which indicates the robustness against the noisy input. Summarizing above, we can assess the robustness against noisy perturbation by visualizing the change ratio of the error-curve surface.

4 Concluding Remarks

We compared the linear stochastic estimation (LSE) and the multi-layer perceptron (MLP) to explore fundamental differences between them by focusing on the noise robustness. The models estimated high-order proper orthogonal decomposition (POD) coefficients from low-order counterparts of a flow around a two-dimensional cylinder. The nonlinear MLP can estimate the coefficients properly from 1st-order input, while the LSE and linear MLP required up to 3rd-order input for estimation. In addition, we found that the LSE is more sensitive to the noise than MLP. To clarify the cause of the difference in response to noise, the error surface visualization was performed. The error surface and the optimized weight of LSE were further different from that of the linear MLP, because of the difference of optimization method. Moreover, the noise addition to the input greatly deformed the error surface of the LSE, and this drastically pushed up the error at the optimization point, comparing to the linear MLP. We found that the robustness against noisy data can be assessed by visualizing the change ratio of the error-curve surface.

References

R. J. Adrian and P. Moin. Stochastic estimation of organized turbulent structure: homogeneous shear flow. *J. Fluid Mech.*, 190, 1988.

S. L. Brunton and J. N. Kutz. *Data-driven science and engineering: Machine learning, dynamical systems, and control*. Cambridge University Press, 2019.

H. Kor, M. Badri Ghomizad, and K. Fukagata. A unified interpolation stencil for ghost-cell immersed boundary method for flow around complex geometries. *J. Fluid Sci. Technol.*, 12(1): JFST0011, 2017.

J.-Ch. Loiseau, S. L. Brunton, and B. R. Noack. From the POD-Galerkin method to sparse manifold models. available on ResearchGate, 2018. doi: 10.13140/RG.2.2.27965.31201.

D. E. Rumelhart, G. E. Hinton, and R. J. Williams. Learning representations by back-propagation errors. *Nature*, 322:533–536, 1986.