**Learning Interpretable Representation of Koopman Operator for Non-linear Dynamics**

Deepak Kumar,a Vinayak Dixit,a Manojkumar Ramteke a,b and Hariprasad Kodamana a,b,\*

a Department of Chemical Engineering, Indian Institute of Technology Delhi, New Delhi, 110016,  
India

b Yardi School of Artificial Intelligence, Indian Institute of Technology Delhi, New Delhi, 110016,  
India

\*kodamana@chemical.iitd.ac.in

Abstract

In this study, we introduce an innovative approach for extracting linear representations from nonlinear dynamical systems lifted to a higher-dimensional space. The Koopman Operator (KO), a powerful tool used to linearize nonlinear dynamics, forms the cornerstone of our approach. However, the analytical identification of KO often presents significant challenges, leading to the reliance on deep learning techniques. These techniques, while powerful, suffer from a lack of transparency due to their 'black-box' nature, thus limiting their broader applicability in scientific domains.

To address this limitation, we employ the Equation Learner (EQL) network in place of conventional neural networks. The EQL, known for its interpretability, is adept at producing explicit mathematical equations that represent the latent embeddings of the system used to convert state variables to lifted space and bring them back to original space. It not only preserves the predictive strength inherent in deep learning models but also infuses the process with the clarity and interpretability typically associated with symbolic regression.

As a practical application of our methodology, we have applied this algorithm to a Continuous Stirred-Tank Reactor (CSTR) example. This implementation successfully linearizes the nonlinear dynamics of the CSTR, demonstrating the efficacy of our approach in a real-world scenario.

**Keywords**: Koopman Operator, Equation Learner, Data-driven Discovery

* 1. Introduction

The landscape of scientific exploration and understanding, especially in the realm of complex dynamical systems, is undergoing a significant transformation. Linearizing non-linear dynamics of chemical processes is preferred due to the simplicity and ease of analysis that linear models offer. These models are more computationally efficient and easier to use for real-time process control and optimization. Additionally, linear control theories, such as PID control, are well-established and more straightforward to implement, making them a practical choice for consistent process performance and stability.

The Koopman operator (KO) provides a powerful method for linearizing complex dynamics by transforming a non-linear dynamical system into an infinite-dimensional linear system [Koopman (1931)]. It does this by acting on the space of observable functions of the state, rather than directly on the state itself. This approach allows the dynamics of the system to be described linearly, even though the system itself may be inherently non-linear. By doing so, KO facilitates the use of linear analysis techniques on complex, non-linear systems, making them more accessible and easier to understand and predict.

The analytical identification of KO, crucial for linearizing complex dynamics, presents a substantial challenge due to its infinite-dimensional nature. This complexity makes deriving an exact analytical representation of KO for complex systems a daunting task. However, Autoencoders, renowned for their ability to compress and encode high-dimensional data into a more manageable form, can approximate KO by transforming the intricate, non-linear state space of a dynamical system into a higher-dimensional, linearly analyzable space. This approach enables the application of linear techniques to systems that are fundamentally non-linear.

Despite this innovative application, autoencoders are often viewed as "black boxes," presenting a significant limitation in terms of interpretability, posing a challenge in scenarios where understanding the underlying process is crucial. This lack of transparency and interpretability is particularly problematic in fields where decision-making relies on a clear understanding of the system's behavior, such as in safety-critical applications like chemical engineering. Thus, there is an increasing need for advancements that balance predictive accuracy with transparency and interpretability in these models.

Integrating Equation Learner (EQL) networks into the encoder and decoder components of an autoencoder used for estimating Koopman observables represents a significant advancement in enhancing interpretability [Martius and Lampert (2016)]. EQL networks, known for their capacity in symbolic regression, generate explicit mathematical equations rather than just numerical outputs. This attribute is key to understanding the inner workings of the model. Incorporating EQLs into encoder and decoder transforms these processes into equation-driven operations, thus making the approach of approximating Koopman observables transparent and interpretable.

The use of EQLs means that the Koopman observables are expressed through clear mathematical relationships, elucidating how different variables and states interact within the system's dynamics. This level of clarity is essential, particularly in complex systems analysis, as it allows for a deeper understanding of the interconnections and dependencies within the system. Furthermore, the explicit equations provided by EQLs facilitate thorough analysis and validation, aligning the model's outputs with established physical laws and theoretical principles. Such validation is critical in fields where accuracy needs to be complemented with theoretical conformity.

Moreover, the interpretability and transparency offered by EQL-based autoencoders significantly enhance the trust and usability of these models in practical scenarios. Practitioners and researchers are more likely to adopt models that provide not just accurate predictions but also a comprehensible rationale behind these predictions. This approach combines the predictive strength of machine learning with the much-needed element of interpretability, making these models particularly valuable in applications where understanding the underlying dynamics is as crucial as the predictive accuracy itself.

* 1. Methodology
     1. Koopman Theory

Koopman theory states that any nonlinear system can be evolved in a linear system in infinite-dimensional space using KO. This provides an elegant approach to analyse and control nonlinear dynamical systems linearly [Lusch et al. (2018)]. Among the current techniques to estimate the KO, one of the most prevalent is extended dynamic mode decomposition, where the operator is identified from a dataset that is linked either spatially or temporally, depending on the description of the system [Proctor et al. (2016)]. In this methodology, the observables of the system are transformed from their original vector space to a different one which has a higher dimension than the native state [Schmid (2010]. This transformation is deliberately made a nonlinear one and the set of these nonlinear mappings is called the dictionary. In the new higher dimensional vector space, it is theorized that the states evolve linearly.

Let us look at a nonlinear dynamical system in discrete-time domain, which can be represented by

|  |  |
| --- | --- |
|  | (1) |

where , is a continuously differentiable mapping. Now, let the elements of dictionary be represented by where, , then according to Koopman theory we get,

|  |  |
| --- | --- |
|  | (2) |

where , is a higher dimensional space, that represents the lifted space. The collection set of all such is called dictionary denoted by D.

|  |  |
| --- | --- |
|  | (3) |

The next step is to create two vectors at different time steps or states, and , which represent the evolution of the system at a state and one step after it.

|  |  |
| --- | --- |
|  | (4) |

where is the state variable at time step k.

Then we can write,

|  |  |
| --- | --- |
|  | (5) |

The KO is then estimated using extended dynamic mode decomposition by the optimization of following expression as problem:

|  |  |
| --- | --- |
|  | (6) |

* + 1. Equation Learner (EQL) Network

Neural networks are, by definition, black box model that find an unknown relation between the input and the output data set provided. If the data is sufficient enough, it is expected to generalize the results to a new data set provided, which is extrapolation. But black box models are not useful as they don't provide any insight or understanding of the system which is desired in industries for process safety and control. For the purpose of learning approximate analytical representations from the neural network, we implement an EQL framework [Martius and Lampert (2016]. A feedforward layered deep neural network was implemented. The initial layers consisted of a linear mapping to a vector which is further transformed using activation functions and mathematical combinations.

The linear mapping of the layer can be defined as,

|  |  |
| --- | --- |
|  | (7) |

Further, the elementary function is operated on element of .

|  |  |
| --- | --- |
|  | (8) |

where,

Furthermore, elementary operations like multiply, divide, etc. are applied among selected elements of as shown in Figure 1.

The parameters of the network that are to be optimized using training can be denoted by , where

|  |  |
| --- | --- |
|  | (9) |

which could be found by minimizing the Mean Squared Error Loss function combined with a regularisation function.

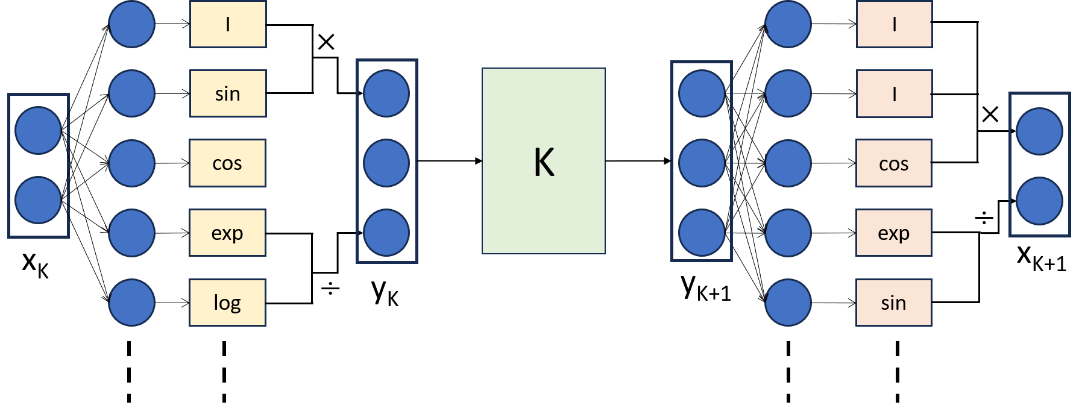


Figure 1: Illustration of integration of Koopman operator and equation learner framework

* + - 1. Interpretable Koopman Operator Estimation

We propose a new method to discover using EQL as shown in Figure 1. Encoder EQL (J) network is used to encode the original input to lifted inputs which is linearly mapped to its next state using Koopman operator (). Furtehr, we use a decoder EQL (L) network to get back the next original inputs from . The final loss value is a combination of following losses:

|  |  |
| --- | --- |
|  | (10) |
|  | (11) |
|  | (12) |
|  | (13) |

In these equations, denotes the squared Euclidean norm.

The regularization is designed to be less sensitive to outliers than the norm and to promote sparsity more than the norm. [Kim et al. (2020)] The is defined as a piecewise function that changes its behavior based on the magnitude of the parameter relative to a threshold . In this case study, the regularization is found to work best at .

* 1. Results and Discussion

The proposed framework's performance in linearizing the dynamics of reversible reaction in a Continuous Stirred Tank Reactor (CSTR) under isothermal conditions was evaluated through its ability to predict the concentrations of reactants A, , and B, over time. The encoder part of the network (J) lifts the 2 inputs to 5 inputs and decoder part provides us next states in original space.

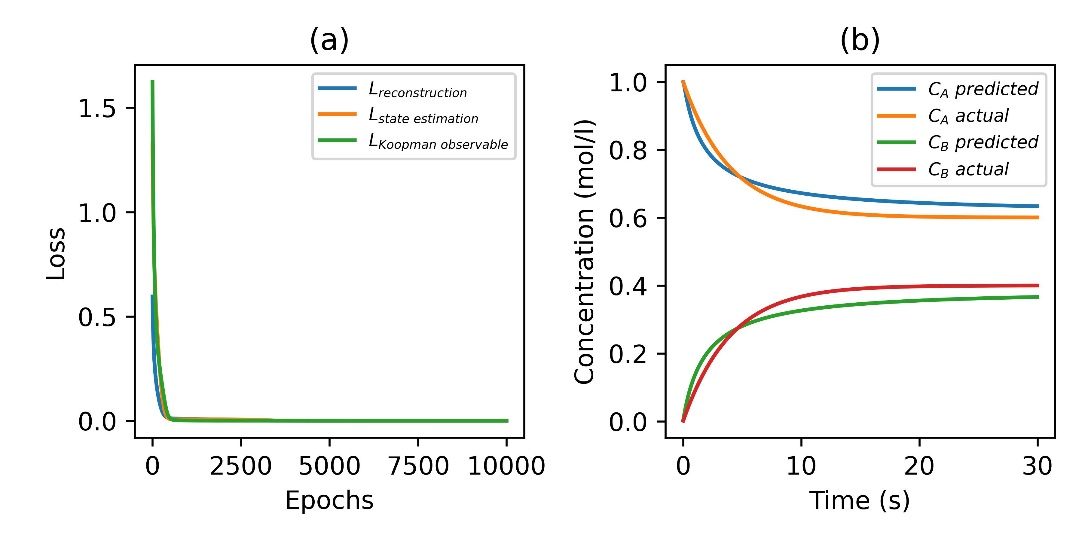


Figure 2: (a) Various training losses vs epochs and (b) State estimation using initial condition and recursive prediction on the training data. Note that training was done for one time step ahead prediction whereas the plots show prediction based on initial conditions only

As shown in Figure 2(a), the convergence of loss metrics, encompassing , , and losses, was rapid and stabilized at near-zero values early in the training process. This fast convergence suggested a strong initial learning phase, attributed to the effective architecture and optimization via gradient descent. The stability of these losses at minimal values without overfitting corroborated the architecture's capability to generalize well to the dynamics of the CSTR system.

Complementing the loss metrics, as shown in Figure 2(b), the predictions of reactant concentrations, and , showed a remarkable alignment with the actual data. It also shows the recursive prediction capabilities of the network. Although trained for one-step-ahead predictions, the graph reveals the model's ability to predict the time evolution of and by recursively applying its learned dynamics, starting from the initial conditions and proceeding without further reference to actual data points. The close alignment between the recursive predictions and the actual data underscores the model's robust understanding of the reaction kinetics, capturing the depletion and formation of the reactants over time.

The application of the regularization played a pivotal role in the network's performance, ensuring a balance between model complexity and predictive precision. This regularization likely prevented overfitting, a common challenge in neural network training.

Also, the use of EQL networks in the encoder and decoder segments afforded complete interpretability of the learned transformations. This transparency is a substantial advantage, as it not only validates the internal workings of the model but also provides explicit mathematical formulations of the observable functions governing the reaction kinetics. These formulations offer a deeper analytical insight into the reaction process, enabling a clear understanding of how the model arrives at its predictions.

Our approach, focusing on interpretable, data-driven estimation of the KO via the EQL network, inherently differs from conventional deep neural network (DNN) methods. The core of our methodology is interpretability, a feature typically not prioritized in other linearization techniques. This unique focus makes direct comparisons with other methods less relevant, as they do not share this central aspect of interpretability, crucial in our research for understanding complex system dynamics. Our research aims to set a new benchmark in the field by demonstrating how interpretability can be integrated into the estimation of the KO, offering clear insights into the dynamics of complex systems.

* 1. Conclusion and Future Work

The combination of a well-designed neural network architecture with EQL networks has demonstrated significant promise in accurately capturing and predicting complex chemical reaction dynamics, offering both predictive power and valuable interpretive insights. Future work should validate the model's effectiveness beyond the training set to ensure that it can generalize to different initial conditions and system perturbations. We propose that the application of our linearized models to specific control contexts represents a significant and promising avenue for future research. This could potentially involve the integration of our methodology into various control scenarios to demonstrate its practical utility and effectiveness. The ability to generate explicit Koopman observables and their inverses also opens up new avenues for leveraging the model in educational settings, where understanding the mechanics of chemical processes is paramount.

References

Schmid, P. J. (2010). Dynamic mode decomposition of numerical and experimental data. Journal of fluid mechanics, 656, 5-28.

Lusch, B., Kutz, J. N., & Brunton, S. L. (2018). Deep learning for universal linear embeddings of nonlinear dynamics. Nature communications, 9(1), 4950.

Li, Q., Dietrich, F., Bollt, E. M., & Kevrekidis, I. G. (2017). Extended dynamic mode decomposition with dictionary learning: A data-driven adaptive spectral decomposition of the Koopman operator. Chaos: An Interdisciplinary Journal of Nonlinear Science, 27(10).

Proctor, J. L., Brunton, S. L., & Kutz, J. N. (2016). Dynamic mode decomposition with control. SIAM Journal on Applied Dynamical Systems, 15(1), 142-161.

Proctor, J. L., Brunton, S. L., & Kutz, J. N. (2018). Generalizing Koopman theory to allow for inputs and control. SIAM Journal on Applied Dynamical Systems, 17(1), 909-930.

Martius, G., & Lampert, C. H. (2016). Extrapolation and learning equations. arXiv preprint arXiv:1610.02995.

Xiao, Y., Zhang, X., Xu, X., Liu, X., & Liu, J. (2020). A deep learning framework based on Koopman operator for data-driven modeling of vehicle dynamics. arXiv preprint arXiv:2007.02219.

Koopman, B. O. (1931). Hamiltonian systems and transformation in Hilbert space. Proceedings of the National Academy of Sciences, 17(5), 315-318.

Kim, S., Lu, P. Y., Mukherjee, S., Gilbert, M., Jing, L., Čeperić, V., & Soljačić, M. (2020). Integration of neural network-based symbolic regression in deep learning for scientific discovery. IEEE transactions on neural networks and learning systems, 32(9), 4166-4177.