PyKoopman: A Python Package for Data-Driven Approximation of the Koopman Operator

07 May 2023

Summary

`PyKoopman` is a Python package for the data-driven approximation of the Koopman operator in the dynamical systems. The Koopman operator has emerged as a principled linear embedding of nonlinear dynamics and facilitates the prediction, estimation, and control of strongly nonlinear dynamics using linear systems theory. In particular, `PyKoopman` provides tools for data-driven system identification for unforced and actuated systems that build on the equation-free dynamic mode decomposition (DMD) (Schmid 2010) and its variants. In this work, we provide a brief description of the mathematical underpinnings of the Koopman operator, an overview and demonstration of the features implemented in `PyKoopman` (with code examples), practical advice for users, and a list of potential extensions to `PyKoopman`. Software is available at https://github.com/dynamicslab/pyKoopman.

Statement of need

Engineers have long employed linearization to bridge the gap between closed-form linear theory and the intricate nonlinear reality (Ljung 2010; Wright, Nocedal, and others 1999). Local linearization, using low-order Taylor expansion, has been prevalent in system identification (Ljung 2010), optimization (Wright, Nocedal, and others 1999), and numerous other fields to render problems more manageable. However, the current expansion of available measurement data, coupled with challenges posed by increasingly complex systems that resist first-principle-based analysis, has spurred the adoption of data-driven modeling techniques. These methods aim to extract a linear embedding from nonlinear dynamics (Brunton and Kutz 2022).

Within the diverse landscape of data-driven modeling approaches for unknown systems, the Koopman operator theory has garnered considerable attention in recent years (Budišić, Mohr, and Mezić 2012; Mezić 2013; Williams, Kevrekidis, and Rowley 2015; Klus et al. 2018; Li et al. 2017; Brunton et al. 2017). These approaches cover not only linear methods (Nelles 2020; Ljung 2010) and dynamic mode decomposition (DMD) (Schmid 2010; Kutz et al. 2016), but also more advanced techniques such as nonlinear autoregressive algorithms (Akaike 1969; Billings 2013), neural networks (Long et al. 2018; Yang, Zhang, and Karniadakis 2020; Wehmeyer and Noé 2018; Mardt et al. 2018; Vlachas et al. 2018; Pathak et al. 2018; Lu et al. 2021; M Raissi, Perdikaris, and Karniadakis 2019; Champion et al. 2019; Raissi, Yazdani, and Karniadakis 2020), Gaussian process regression (Maziar Raissi, Babaee, and Karniadakis 2019), operator inference and reduced-order modeling (Benner, Gugercin, and Willcox 2015; Peherstorfer and Willcox 2016; Qian et al. 2020), among others (Giannakis and Majda 2012; Yair et al. 2017; Bongard and Lipson 2007; Schmidt and Lipson 2009; Daniels and Nemenman 2015; Brunton, Proctor, and Kutz 2016; Rudy et al. 2017). The Koopman operator occupies a unique position within data-driven modeling techniques due to its distinct objective of achieving global linearization of nonlinear dynamics based on data. This approach allows for the application of closed-form, convergence-guaranteed methods from linear system theory to general nonlinear dynamics. In contrast, most other algorithms primarily concentrate on learning end-to-end and/or black-box mappings from parameters to solutions, which sets the Koopman operator apart from its counterparts. To fully harness the potential of data-driven global linearization across a broad array of scientific and engineering disciplines, it is essential to develop tools that automate state-of-the-art Koopman operator algorithms.

As a result, the `PyKoopman` is developed as a Python package for approximating the Koopman operator associated with natural and actuated dynamical systems from data. In particular, `PyKoopman` provides tools for designing the observables (e.g., functions of system state) and inferring a finite-dimensional linear operator that governs the observables. These two steps can be either performed sequentially (Williams, Rowley, and Kevrekidis 2015; Williams, Kevrekidis, and Rowley 2015) or combined as in the case of more recent neural network models (Lusch, Kutz, and Brunton 2018; Otto and Rowley 2019). After a linear embedding is discovered from the data, one can leverage the linearity of the lifted dynamical system for better interpretability (Pan, Arnold-Medabalimi, and Duraisamy 2021) or near-optimal observer (Surana and Banaszuk 2016) or controller on the original nonlinear system (Korda and Mezić 2020; Mauroy, Susuki, and Mezić 2020; Kaiser, Kutz, and Brunton 2021).

New features

The core component of the `PyKoopman` package is the `Koopman` model class. To make this package accessible to a broader user base, this class is implemented as a `scikit-learn` estimator. The external package dependencies are illustrated in Fig. 2. Additionally, users can create sophisticated pipelines for hyperparameter tuning and model selection by integrating `pyKoopman` with `scikit-learn`.

As illustrated in Fig. 3, `PyKoopman` is designed to lift nonlinear dynamics into a linear system with linear actuation. Specifically, our `PyKoopman` implementation involves two major steps:

- `observables`: the nonlinear observables used to lift $\mathbf{x}$ to $\mathbf{z}$, and reconstruct $\mathbf{x}$ from $\mathbf{z}$;
- `regression`: the regression used to find the optimal $\mathbf{A}$.

Additionally, we have a `differentiation` module that evaluates the time derivative from a trajectory and the `analytics` module for sparsifying arbitrary approximations of the Koopman operator.

At the time of writing, we have the following features implemented:

- Observable library for lifting the state $\mathbf{x}$ into the observable space
  - Identity (for DMD/DMDc or in case users want to compute observables themselves): `Identity`
  - Multivariate polynomials: `Polynomial`
  - Time delay coordinates: `TimeDelay`
  - Radial basis functions: `RadialBasisFunctions`
  - Random Fourier features: `RandomFourierFeatures`
  - Custom library (defined by user-supplied functions): `CustomObservables`
  - Concatenation of observables: `ConcatObservables`
- System identification method for performing regression
  - Dynamic mode decomposition (Schmid 2010; Rowley et al. 2009): `PyDMDRegressor`
  - Dynamic mode decomposition with control (Proctor, Brunton, and Kutz 2016): `DMDc`
  - Extended dynamic mode decomposition (Williams, Kevrekidis, and Rowley 2015): `EDMD`
  - Extended dynamic mode decomposition with control (Korda and Mezić 2020): `EDMDc`
  - Kernel dynamic mode decomposition (Williams, Rowley, and Kevrekidis 2015): `KDMD`
  - Hankel DMD (Brunton, Proctor, and Kutz 2016): `HDMD`
  - Hankel DMD with control: `HDMDc`
  - Neural Network DMD (Pan and Duraisamy 2020; Otto and Rowley 2019; Lusch, Kutz, and Brunton 2018): `NNDMD`
- Sparse construction of Koopman invariant subspace
  - Multi-task learning based on linearity consistency(Pan, Arnold-Medabalimi, and Duraisamy 2021): `ModesSelectionPAD21`
- Numerical differentiation for computing $\dot{\mathbf{X}}$ from $\mathbf{X}$
  - Finite difference: `FiniteDifference`
  - 4th order central finite difference: `Derivative(kind="finite_difference")`
  - Savitzky-Golay with cubic polynomials: `Derivative(kind="savitzky-golay")`

- Spectral derivative: `Derivative(kind="spectral")`
  - Spline derivative: `Derivative(kind="spline")`
  - Regularized total variation derivative: `Derivative(kind="trend_filtered")`
- Common toy dynamics
  - Discrete-time random, stable, linear state-space model: `drss`
  - Van del Pol oscillator: `vdp_osc`
  - Lorenz system: `lorenz`
  - Two-dimensional linear dynamics: `Linear2Ddynamics`
  - Linear dynamics on a torus: `torus_dynamics`
  - Forced Duffing Oscillator: `forced_duffing`
  - Cubic-quintic Ginzburg-Landau equation: `cqgle`
  - Kuramoto-Sivashinsky equation:`ks`
  - Nonlinear Schrodinger equation: `nls`
  - Viscous Burgers equation: `vbe`
- Validation routines for consistency checks

Conclusion

Our goal of the `PyKoopman` package is to provide a central hub for education, application and research development of learning algorithms for Koopman operator. The `PyKoopman` package is aimed at researchers and practitioners alike, enabling anyone with access to discover linear embeddings of nonlinear systems from data. The package is designed to be accessible to users with basic knowledge of linear systems, adhering to `scikit-learn` standards, while also being modular for more advanced users. We hope that researchers and practioners will use `PyKoopman` as a platform for algorithms developement and applications of linear embedding.

Acknowledgments

References

Akaike, Hirotugu. 1969. "Fitting Autoregressive Models for Prediction." *Annals of the Institute of Statistical Mathematics* 21 (1): 243–47.

Benner, Peter, Serkan Gugercin, and Karen Willcox. 2015. "A Survey of Projection-Based Model Reduction Methods for Parametric Dynamical Systems." *SIAM Review* 57 (4): 483–531.

Billings, Stephen A. 2013. *Nonlinear System Identification: NARMAX Methods in the Time, Frequency, and Spatio-Temporal Domains*. John Wiley & Sons.

Bongard, Josh, and Hod Lipson. 2007. "Automated Reverse Engineering of Nonlinear Dynamical Systems." *Proceedings of the National Academy of*

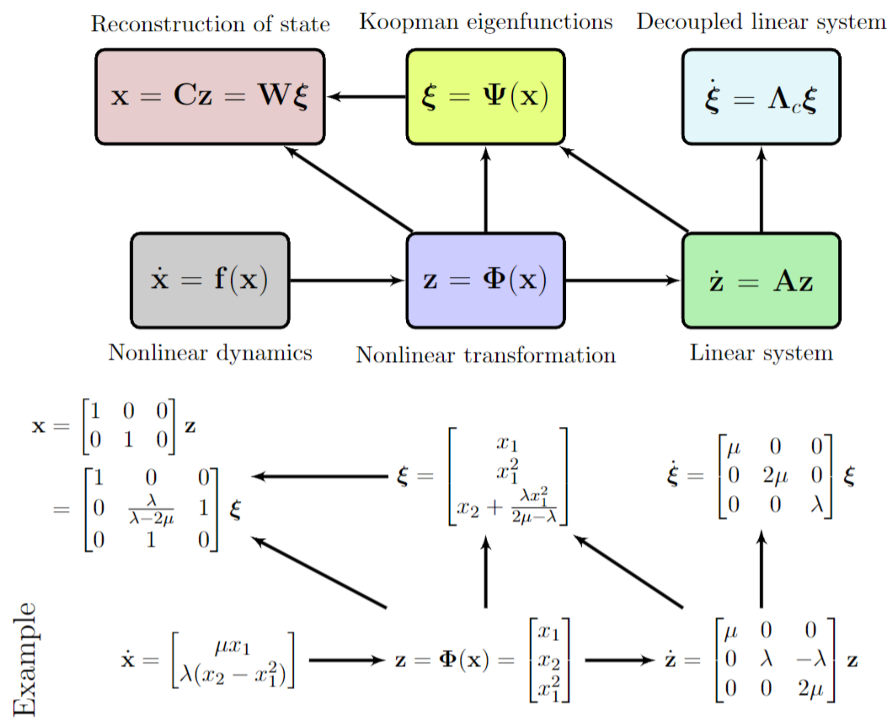

Figure 1: Lifting of the state $\mathbf{x}$ of the continuous autonomous dynamical system into a new coordinate system, in which the original nonlinear dynamics become linear and are easier to handle. One can also linearly reconstruct the state $\mathbf{x}$ from the new coordinate system. This is facilitated with `PyKoopman` in a data-driven manner.

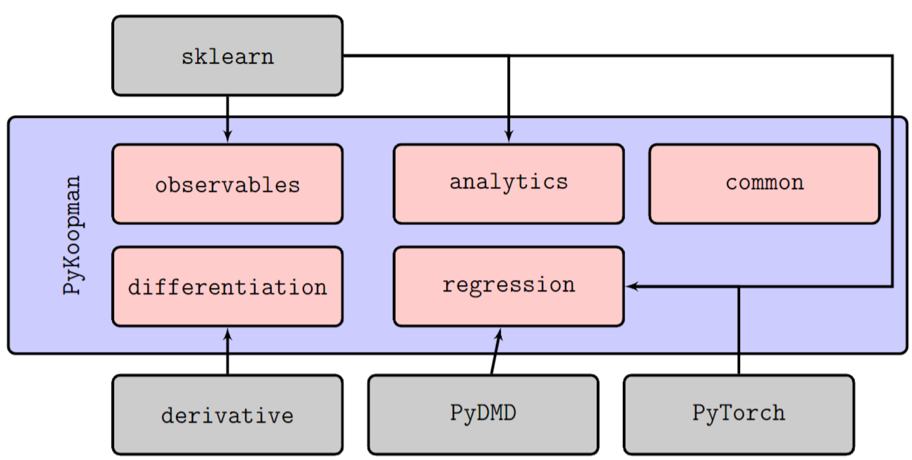

Figure 2: External package dependencies of PyKoopman.

Unforced

$$\mathbf{z}_{k+1} = \mathbf{A}\mathbf{z}_k$$

Controlled

$$\begin{bmatrix} \mathbf{z} \\ \mathbf{u} \end{bmatrix}_{k+1} = \underbrace{\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \cdot & \cdot \end{bmatrix}}_{=\mathbf{K}} \begin{bmatrix} \mathbf{z} \\ \mathbf{u} \end{bmatrix}_{k}$$

Figure 3: {Broad categorization of model types that can be identified with current `PyKoopman`. While the dotted parts (marked with "·") can be simultaneously discovered within the framework, they are typically ignored for control purposes.

Sciences 104 (24): 9943–8.

Brunton, Steven L, Bingni W Brunton, Joshua L Proctor, Eurika Kaiser, and J Nathan Kutz. 2017. "Chaos as an Intermittently Forced Linear System." *Nature Communications* 8 (1): 1–9.

Brunton, Steven L, and J Nathan Kutz. 2022. *Data-Driven Science and Engineering: Machine Learning, Dynamical Systems, and Control*. Cambridge University Press.

Brunton, Steven L., Joshua L. Proctor, and J. Nathan Kutz. 2016. "Discovering Governing Equations from Data by Sparse Identification of Nonlinear Dynamical Systems." *Proceedings of the National Academy of Sciences* 113 (15): 3932–7. https://doi.org/10.1073/pnas.1517384113.

Budišić, Marko, Ryan Mohr, and Igor Mezić. 2012. "Applied Koopmanism." *Chaos: An Interdisciplinary Journal of Nonlinear Science* 22 (4): 047510.

Champion, Kathleen, Bethany Lusch, J Nathan Kutz, and Steven L Brunton. 2019. "Data-Driven Discovery of Coordinates and Governing Equations." *Proceedings of the National Academy of Sciences* 116 (45): 22445–51.

Daniels, Bryan C, and Ilya Nemenman. 2015. "Automated Adaptive Inference of Phenomenological Dynamical Models." *Nature Communications* 6 (1): 1–8.

Giannakis, Dimitrios, and Andrew J Majda. 2012. "Nonlinear Laplacian Spectral Analysis for Time Series with Intermittency and Low-Frequency Variability." *Proceedings of the National Academy of Sciences* 109 (7): 2222–7.

Kaiser, Eurika, J Nathan Kutz, and Steven L Brunton. 2021. "Data-Driven Discovery of Koopman Eigenfunctions for Control." *Machine Learning: Science and Technology* 2 (3): 035023.

Klus, Stefan, Feliks Nüske, Péter Koltai, Hao Wu, Ioannis Kevrekidis, Christof Schütte, and Frank Noé. 2018. "Data-Driven Model Reduction and Transfer Operator Approximation." *Journal of Nonlinear Science* 28 (3): 985–1010.

Korda, Milan, and Igor Mezić. 2020. "Optimal Construction of Koopman Eigenfunctions for Prediction and Control." *IEEE Transactions on Automatic Control* 65 (12): 5114–29.

Kutz, J Nathan, Steven L Brunton, Bingni W Brunton, and Joshua L Proctor. 2016. *Dynamic Mode Decomposition: Data-Driven Modeling of Complex Systems*. SIAM.

Li, Qianxiao, Felix Dietrich, Erik M Bollt, and Ioannis G Kevrekidis. 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): 103111.

Ljung, Lennart. 2010. "Perspectives on System Identification." *Annual Reviews in Control* 34 (1): 1–12.

Long, Zichao, Yiping Lu, Xianzhong Ma, and Bin Dong. 2018. "Pde-Net: Learning Pdes from Data." In *International Conference on Machine Learning*, 3208–16. PMLR.

Lu, Lu, Xuhui Meng, Zhiping Mao, and George Em Karniadakis. 2021. "DeepXDE: A Deep Learning Library for Solving Differential Equations." *SIAM Review* 63 (1): 208–28.

Lusch, Bethany, J Nathan Kutz, and Steven L Brunton. 2018. "Deep Learning for Universal Linear Embeddings of Nonlinear Dynamics." *Nature Communications* 9 (1): 4950.

Mardt, Andreas, Luca Pasquali, Hao Wu, and Frank Noé. 2018. "VAMPnets for Deep Learning of Molecular Kinetics." *Nature Communications* 9 (1): 1–11.

Mauroy, Alexandre, Y Susuki, and I Mezić. 2020. *Koopman Operator in Systems and Control*. Springer.

Mezić, Igor. 2013. "Analysis of Fluid Flows via Spectral Properties of the Koopman Operator." *Annual Review of Fluid Mechanics* 45: 357–78.

Nelles, Oliver. 2020. "Nonlinear Dynamic System Identification." In *Nonlinear System Identification*, 831–91. Springer.

Otto, Samuel E, and Clarence W Rowley. 2019. "Linearly Recurrent Autoencoder Networks for Learning Dynamics." *SIAM Journal on Applied Dynamical Systems* 18 (1): 558–93.

Pan, Shaowu, Nicholas Arnold-Medabalimi, and Karthik Duraisamy. 2021. "Sparsity-Promoting Algorithms for the Discovery of Informative Koopman-Invariant Subspaces." *Journal of Fluid Mechanics* 917: A18.

Pan, Shaowu, and Karthik Duraisamy. 2020. "Physics-Informed Probabilistic Learning of Linear Embeddings of Nonlinear Dynamics with Guaranteed Stability." *SIAM Journal on Applied Dynamical Systems* 19 (1): 480–509.

Pathak, Jaideep, Brian Hunt, Michelle Girvan, Zhixin Lu, and Edward Ott. 2018. "Model-Free Prediction of Large Spatiotemporally Chaotic Systems from Data: A Reservoir Computing Approach." *Physical Review Letters* 120 (2): 024102.

Peherstorfer, Benjamin, and Karen Willcox. 2016. "Data-Driven Operator Inference for Nonintrusive Projection-Based Model Reduction." *Computer Methods in Applied Mechanics and Engineering* 306: 196–215.

Proctor, Joshua L, Steven L Brunton, and J Nathan Kutz. 2016. "Dynamic Mode Decomposition with Control." *SIAM Journal on Applied Dynamical Systems* 15 (1): 142–61.

Qian, Elizabeth, Boris Kramer, Benjamin Peherstorfer, and Karen Willcox. 2020. "Lift & Learn: Physics-Informed Machine Learning for Large-Scale Nonlinear Dynamical Systems." *Physica D: Nonlinear Phenomena* 406: 132401.

Raissi, Maziar, Hessam Babaee, and George Em Karniadakis. 2019. "Parametric Gaussian Process Regression for Big Data." *Computational Mechanics* 64: 409–16.

Raissi, Maziar, Alireza Yazdani, and George Em Karniadakis. 2020. "Hidden Fluid Mechanics: Learning Velocity and Pressure Fields from Flow Visualizations." *Science* 367 (6481): 1026–30.

Raissi, M, P Perdikaris, and GE Karniadakis. 2019. "Physics-Informed Neural Networks: A Deep Learning Framework for Solving Forward and Inverse Problems Involving Nonlinear Partial Differential Equations." *Journal of Computational Physics* 378: 686–707. https://doi.org/10.1016/j.jcp.2018.10.045.

Rowley, Clarence W, Igor Mezić, Shervin Bagheri, Philipp Schlatter, and Dan S Henningson. 2009. "Spectral Analysis of Nonlinear Flows." *Journal of Fluid Mechanics* 641: 115–27.

Rudy, Samuel H, Steven L. Brunton, Joshua L. Proctor, and J. Nathan Kutz. 2017. "Data-Driven Discovery of Partial Differential Equations." *Science Advances* 3 (e1602614). https://doi.org/10.1126/sciadv.1602614.

Schmid, Peter J. 2010. “Dynamic Mode Decomposition of Numerical and Experimental Data.” *Journal of Fluid Mechanics* 656: 5–28.

Schmidt, Michael, and Hod Lipson. 2009. “Distilling Free-Form Natural Laws from Experimental Data.” *Science* 324 (5923): 81–85.

Surana, Amit, and Andrzej Banaszuk. 2016. “Linear Observer Synthesis for Nonlinear Systems Using Koopman Operator Framework.” *IFAC-PapersOnLine* 49 (18): 716–23.

Vlachas, Pantelis R, Wonmin Byeon, Zhong Y Wan, Themistoklis P Sapsis, and Petros Koumoutsakos. 2018. “Data-Driven Forecasting of High-Dimensional Chaotic Systems with Long Short-Term Memory Networks.” *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 474 (2213): 20170844.

Wehmeyer, Christoph, and Frank Noé. 2018. “Time-Lagged Autoencoders: Deep Learning of Slow Collective Variables for Molecular Kinetics.” *The Journal of Chemical Physics* 148 (24): 241703.

Williams, Matthew O, Ioannis G Kevrekidis, and Clarence W Rowley. 2015. “A Data–Driven Approximation of the Koopman Operator: Extending Dynamic Mode Decomposition.” *Journal of Nonlinear Science* 25 (6): 1307–46.

Williams, Matthew O, Clarence W Rowley, and Ioannis G Kevrekidis. 2015. “A Kernel Approach to Data-Driven Koopman Spectral Analysis.” *Journal of Computational Dynamics* 2: 247–65.

Wright, Stephen, Jorge Nocedal, and others. 1999. “Numerical Optimization.” *Springer Science* 35 (67-68): 7.

Yair, Or, Ronen Talmon, Ronald R Coifman, and Ioannis G Kevrekidis. 2017. “Reconstruction of Normal Forms by Learning Informed Observation Geometries from Data.” *Proceedings of the National Academy of Sciences* 114 (38): E7865–E7874.

Yang, Liu, Dongkun Zhang, and George Em Karniadakis. 2020. “Physics-Informed Generative Adversarial Networks for Stochastic Differential Equations.” *SIAM Journal on Scientific Computing* 42 (1): A292–A317.