PySINDy: A Python package for the sparse identification of nonlinear dynamical systems from data

Brian M. de Silva[1], **Kathleen Champion**[1], **Markus Quade**[2], **Jean-Christophe Loiseau**[3], **J. Nathan Kutz**[1], **and Steven L. Brunton**[4, 1]

1 Department of Applied Mathematics, University of Washington **2** Ambrosys GmbH **3** École Nationale Supérieure des Arts et Métiers **4** Department of Mechanical Engineering, University of Washington

DOI: 10.21105/joss.02093

Software

- Review
- Repository
- Archive

Editor: Yuan Tang

Submitted: 11 February 2020
Published: 11 February 2020

Summary

Scientists have long quantified empirical observations by developing mathematical models that characterize the observations, have some measure of interpretability, and are capable of making predictions. Dynamical systems models in particular have been widely used to study, explain, and predict system behavior in a wide range of application areas, with examples ranging from Newton's laws of classical mechanics to the Michaelis-Menten kinetics for modeling enzyme kinetics. While governing laws and equations were traditionally derived by hand, the current growth of available measurement data and resulting emphasis on data-driven modeling motivates algorithmic approaches for model discovery. A number of such approaches have been developed in recent years and have generated widespread interest, including Eureqa (Schmidt & Lipson, 2009), sure independence screening and sparsifying operator (Ouyang, Curtarolo, Ahmetcik, Scheffler, & Ghiringhelli, 2018), and the sparse identification of nonlinear dynamics (SINDy) (Brunton, Proctor, & Kutz, 2016). Maximizing the impact of these model discovery methods requires tools to make them widely accessible to scientists across domains and at various levels of mathematical expertise.

`PySINDy` is a Python package for the discovery of governing dynamical systems models from data. In particular, `PySINDy` provides tools for applying the SINDy approach to model discovery (Brunton et al., 2016). Given data in the form of state measurements $\mathbf{x}(t) \in \mathbb{R}^n$, the SINDy method seeks a function $\mathbf{f}$ such that

$$\frac{d}{dt}\mathbf{x}(t) = \mathbf{f}(\mathbf{x}(t)).$$

SINDy poses this model discovery as a sparse regression problem, wherein relevant terms in $\mathbf{f}$ are selected from a library of candidate functions. Thus, SINDy models balance accuracy and efficiency, resulting in parsimonious models that avoid overfitting while remaining interpretable and generalizable. This approach is straightforward to understand and can be readily customized using different sparse regression algorithms or library functions.

The `PySINDy` package is aimed at researchers and practitioners alike, enabling anyone with access to measurement data to engage in scientific model discovery. The package is designed to be accessible to inexperienced practitioners, while also including options that allow more advanced users to customize it to their needs. A number of popular SINDy variants are implemented, but `PySINDy` is also designed to enable further extensions for research and experimentation. The package follows object-oriented design and is `scikit-learn` compatible.

The SINDy method has been widely applied for model identification in applications such as chemical reaction dynamics (Hoffmann, Fröhner, & Noé, 2019), nonlinear optics (Sorokina,

The Journal of Open Source Software

Sygletos, & Turitsyn, 2016), thermal fluids (Loiseau, 2019), plasma convection (Dam, Brøns, Juul Rasmussen, Naulin, & Hesthaven, 2017), numerical algorithms (Thaler, Paehler, & Adams, 2019), and structural modeling (Lai & Nagarajaiah, 2019). It has also been extended to handle more complex modeling scenarios such as partial differential equations (Rudy, Brunton, Proctor, & Kutz, 2017; Schaeffer, 2017), systems with inputs or control (Kaiser, Kutz, & Brunton, 2018), corrupt or limited data (Schaeffer, Tran, & Ward, 2018; Tran & Ward, 2017), integral formulations (Reinbold, Gurevich, & Grigoriev, 2020; Schaeffer & McCalla, 2017), physical constraints (Loiseau & Brunton, 2018), tensor representations (Gelß, Klus, Eisert, & Schütte, 2019), and stochastic systems (Boninsegna, Nüske, & Clementi, 2018). However, there is not a definitive standard implementation or package for applying SINDy. Versions of SINDy have been implemented within larger projects such as `sparsereg` (Quade, 2018), but no specific implementation has emerged as the most widely adopted and most versions implement only a limited set of features. Researchers have thus typically written their own implementations, resulting in duplicated effort and a lack of standardization. This not only makes it more difficult to apply SINDy to scientific data sets, but also makes it more challenging to benchmark extensions to the method against the original and makes such extensions less accessible to end users. The `PySINDy` package provides a dedicated central codebase where many of the basic SINDy features are implemented, allowing for easy use and standardization. This also makes it straightforward for users to extend the package in a way such that new developments are available to a wider user base.

Features

The core object in the `PySINDy` package is the `SINDy` model class, which is implemented as a `scikit-learn` estimator. This design was chosen to make the package simple to use for a wide user base, as many potential users will be familiar with `scikit-learn`. It also expresses the `SINDy` model object at the appropriate level of abstraction so that users can embed it into more complicated pipelines in `scikit-learn`, such as tools for parameter tuning and model selection.

Applying `SINDy` involves making several modeling decisions, namely: which numerical differentiation method is used, which functions make up the feature library, and which sparse regression algorithm is applied to learn the model. The core `SINDy` object uses a set of default options but can be easily customized using a number of common approaches implemented in `PySINDy`. The package provides a few standard options for numerical differentiation (finite difference and smoothed finite difference), feature libraries (polynomial and Fourier libraries, as well as a class for creating custom libraries), and sparse regression techniques (sequentially thresholded least squares (Brunton et al., 2016), LASSO (Tibshirani, 1996), and sparse relaxed regularized regression (Zheng, Askham, Brunton, Kutz, & Aravkin, 2018)). Users can also create their own differentiation, sparse regression, or feature library objects for further customization.

The software package includes tutorials in the form of Jupyter notebooks. These tutorials demonstrate the usage of various features in the package and reproduce the examples from the original SINDy paper (Brunton et al., 2016).

Acknowledgments

This project is a fork of `sparsereg` (Quade, 2018). SLB acknowledges funding support from the Air Force Office of Scientific Research (AFOSR FA9550-18-1-0200) and the Army Research Office (ARO W911NF-19-1-0045). JNK acknowledges support from the Air Force Office of Scientific Research (AFOSR FA9550-17-1-0329). This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant Number DGE-1256082.

References

Boninsegna, L., Nüske, F., & Clementi, C. (2018). Sparse learning of stochastic dynamical equations. *The Journal of chemical physics*, *148*(24), 241723. doi:10.1063/1.5018409

Brunton, S. L., Proctor, J. L., & Kutz, J. N. (2016). Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proceedings of the National Academy of Sciences*, *113*(15), 3932–3937. doi:10.1073/pnas.1517384113

Dam, M., Brøns, M., Juul Rasmussen, J., Naulin, V., & Hesthaven, J. S. (2017). Sparse identification of a predator-prey system from simulation data of a convection model. *Physics of Plasmas*, *24*(2), 022310. doi:10.1063/1.4977057

Gelß, P., Klus, S., Eisert, J., & Schütte, C. (2019). Multidimensional approximation of nonlinear dynamical systems. *Journal of Computational and Nonlinear Dynamics*, *14*(6). doi:10.1115/1.4043148

Hoffmann, M., Fröhner, C., & Noé, F. (2019). Reactive SINDy: Discovering governing reactions from concentration data. *Journal of Chemical Physics*, *150*(025101). doi:10.1063/1.5066099

Kaiser, E., Kutz, J. N., & Brunton, S. L. (2018). Sparse identification of nonlinear dynamics for model predictive control in the low-data limit. *Proceedings of the Royal Society of London A*, *474*(2219). doi:10.1098/rspa.2018.0335

Lai, Z., & Nagarajaiah, S. (2019). Sparse structural system identification method for nonlinear dynamic systems with hysteresis/inelastic behavior. *Mechanical Systems and Signal Processing*, *117*, 813–842. doi:10.1016/j.ymssp.2018.08.033

Loiseau, J.-C. (2019). Data-driven modeling of the chaotic thermal convection in an annular thermosyphon. *arXiv preprint arXiv:1911.07920*.

Loiseau, J.-C., & Brunton, S. L. (2018). Constrained sparse Galerkin regression. *Journal of Fluid Mechanics*, *838*, 42–67. doi:10.1017/jfm.2017.823

Ouyang, R., Curtarolo, S., Ahmetcik, E., Scheffler, M., & Ghiringhelli, L. M. (2018). SISSO: A compressed-sensing method for identifying the best low-dimensional descriptor in an immensity of offered candidates. *Physical Review Materials*, *2*(8), 083802. doi:10.1103/physrevmaterials.2.083802

Quade, M. (2018, February). Sparsereg - collection of modern sparse regression algorithms. doi:10.5281/zenodo.1173754

Reinbold, P. A., Gurevich, D. R., & Grigoriev, R. O. (2020). Using noisy or incomplete data to discover models of spatiotemporal dynamics. *Physical Review E*, *101*(1), 010203. doi:10.1103/physreve.101.010203

Rudy, S. H., Brunton, S. L., Proctor, J. L., & Kutz, J. N. (2017). Data-driven discovery of partial differential equations. *Science Advances*, *3*(e1602614). doi:10.1126/sciadv.1602614

Schaeffer, H. (2017). Learning partial differential equations via data discovery and sparse optimization. In *Proceedings of the royal society a* (Vol. 473, p. 20160446). The Royal Society. doi:10.1098/rspa.2016.0446

Schaeffer, H., & McCalla, S. G. (2017). Sparse model selection via integral terms. *Physical Review E*, *96*(2), 023302. doi:10.1103/physreve.96.023302

Schaeffer, H., Tran, G., & Ward, R. (2018). Extracting sparse high-dimensional dynamics from limited data. *SIAM Journal on Applied Mathematics*, *78*(6), 3279–3295. doi:10.1137/18m116798x

Schmidt, M., & Lipson, H. (2009). Distilling free-form natural laws from experimental data. *Science*, *324*(5923), 81–85. doi:10.1126/science.1165893

Sorokina, M., Sygletos, S., & Turitsyn, S. (2016). Sparse identification for nonlinear optical communication systems: SINO method. *Optics express*, *24*(26), 30433–30443. doi:10.1364/oe.24.030433

Thaler, S., Paehler, L., & Adams, N. A. (2019). Sparse identification of truncation errors. *Journal of Computational Physics*, *397*, 108851. doi:10.1016/j.jcp.2019.07.049

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, *58*(1), 267–288. doi:10.1111/j.2517-6161.1996.tb02080.x

Tran, G., & Ward, R. (2017). Exact recovery of chaotic systems from highly corrupted data. *Multiscale Modeling & Simulation*, *15*(3), 1108–1129. doi:10.1137/16m1086637

Zheng, P., Askham, T., Brunton, S. L., Kutz, J. N., & Aravkin, A. Y. (2018). A unified framework for sparse relaxed regularized regression: SR3. *IEEE Access*, *7*, 1404–1423. doi:10.1109/access.2018.2886528