IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2023.1120000

Learning Nonlinear Dynamics Using Kalman Smoothing

JACOB M. STEVENS-HAAS[1], YASH BHANGALE[2], J. Nathan Kutz[1 3], and Aleksandr Aravkin[1]
[1] Department of Applied Mathematics, University of Washington, Seattle, WA 98195-3925 USA
[2]Department of Mechanical Engineering, University of Washington
[3]Department of Electrical and Computer Engineering, University of Washington
Corresponding author: Jacob M. Stevens-Haas (e-mail: jacob.stevens.haas@gmail.com).

This work was supported in part by the US National Science Foundation (NSF) AI Institute for Dynamical Systems (dynamicsai.org), grant 2112085 and by the US Department of Veterans Affairs under the Post-9/11 GI Bill

ABSTRACT Identifying Ordinary Differential Equations (ODEs) from measurement data requires both fitting the dynamics and assimilating, either implicitly or explicitly, the measurement data. The *Sparse Identification of Nonlinear Dynamics* (SINDy) method involves a derivative estimation (and optionally, smoothing) step and a sparse regression step on a library of candidate ODE terms. Kalman smoothing is a classical framework for assimilating the measurement data with known noise statistics. Previously, derivatives in SINDy and its python package, pysindy, had been estimated by finite difference, L1 total variation minimization, or local filters like Savitzky-Golay. In contrast, Kalman allows discovering ODEs that best recreate the essential dynamics in simulation, even in cases when it does not perform as well at recovering coefficients, as measured by their F1 score and mean absolute error. We have incorporated Kalman smoothing, along with hyperparameter optimization, into the existing pysindy architecture, allowing for rapid adoption of the method. Numerical experiments on a number of dynamical systems show Kalman smoothing to be the most amenable to parameter selection and best at preserving problem structure in the presence of noise.

INDEX TERMS Dynamical systems, machine learning, sparse regression, optimization, Kalman smoothing, SINDy, differential equations.

I. INTRODUCTION

The method of *Sparse Identification of Nonlinear Dynamics* (SINDy) [Brunton et al., 2016, Brunton and Kutz, 2022] seeks to discover a differential or partial differential equation governing an arbitrary, temporally measured system. The method takes as input some coordinate measurements over time, such as angles between molecular bonds [Boninsegna et al., 2018] or a spatial field, such as wave heights [Rudy et al., 2017], and returns the best ordinary or partial differential equation (ODE or PDE) from a library of candidate terms. However, the method struggles to accommodate significant measurement noise, which is typical of real-world systems. On the other hand, Kalman theory [Kalman, 1960, Kalman and Bucy, 1961] has a half-century history of assimilating measurement noise to smooth a trajectory, with well-studied and rigorously characterized noise properties [Welch et al., 1995]. We integrate the mature and well-established theory of Kalman with the emerging SINDy technology and combine with generalized cross validation (GCV) parameter selection for systematic practical applications. Our Kalman SINDy architecture is shown to be competitive with other combinations of data smoothing and system identification techniques, and has a significant advantage in preservation of problem structure and ease of parameter selection.

Model discovery methods are emerging as a critical component in data-driven engineering design and scientific discovery. Enabled by advancements in computational power, optimization schemes, and machine learning algorithms, such techniques are revolutionizing what can be achieved from sensor measurements deployed in a given system. Of interest here is the discovery of dynamic models, which can be constructed from a diversity of techniques, including simple regression techniques such as the *dynamic mode decomposition* (DMD) [Kutz et al., 2016, Ichinaga et al., 2024] to neural networks such as *physics-informed neural networks* (PINNs) [Raissi et al., 2019]. In such models, the objective is to construct a proxy model for the observed measurements which an be used to characterize and reconstruct solutions.

While DMD provides an interpretable model in terms of a modal decomposition, most neural network architectures remain black-box without a clear view of the underlying dynamical processes. Although the number of techniques available are beyond the scope of this paper to review [Cuomo et al., 2022, North et al., 2023], SINDy is perhaps the leading data-driven model discovery method for interpretable and/or explainable dynamic models as it looks to infer the governing equations underlying the observed data. As such, it discovers relationships between spatial and/or temporal derivatives, which is the underlying mathematical representation of physics and engineering based systems since the time of Newton.

The SINDy regression architecture seeks to robustly establish relationships between derivatives. Emerging from [Brunton et al., 2016, Brunton and Kutz, 2022], all variants aim to discover a sparse symbolic representation of an autonomous or controlled system, $\dot{x} = f(x)$. A diversity of methodological innovations have been introduced into the SINDy discovery framework to make it robust and stable, including the *weak form* optimization by Messenger and Bortz [Messenger and Bortz, 2021a,b]. This approach solves the sparse regression problem after integrating the data over random control volumes, providing a dramatic improvement to the noise robustness of the algorithm. Weak form optimization may be thought of as a generalization of the integral SINDy [Schaeffer and Mccalla, 2017] to PDE-FIND. Further improvements to noise robustness and limited data may be obtained through ensembling techniques [Fasel et al., 2022], which use robust statistical bagging to learn inclusion probabilities for the sparse terms ξ, similar to Bayesian inference [Gao and Kutz, 2022, Gao et al., 2023, Hirsh et al., 2022]. Many methodological innovations are integrated in the open-source PySINDy software library [Kaptanoglu et al., 2022], reducing the barrier to entry when applying these methods to new problems. Additional techniques for learning dynamics from data include PDE-NET [Long et al., 2019, 2018] and the Bayesian PDE discovery from data [Atkinson, 2020]. Symbolic learning has also been developed, including symbolic learning on graph neural networks [Cranmer et al., 2019, 2020, Sanchez-Gonzalez et al., 2020].

Kalman smoothing, which this paper integrates with SINDy, has a long history of assimilating measurement data in time series. From its debut in Kalman [1960], Kalman and Bucy [1961], engineering practice and design have used it for control and prediction across the real world, e.g. in radar systems, econometric variables, weather prediction, and more. The family of Kalman methods encompasses both smoothing, after-the-fact techniques, and filtering, real-time updates, that derive from the same assumptions for distributions. The Kalman smoother can be considered as a best-fit Euler update, the maximum likelihood estimator of Brownian motion, or as the best linear fit of an unknown system. The best fit/maximum likelihood view extends the classic Kalman updates to a rich family of efficient generalized Kalman smooth algorithms for signals corrupted by outliers, nonlinear models, constraints through side information, and a myriad of other applications, see Aravkin et al. [2017, 2012], Jonker et al. [2019]. In the simplest invocation, the Kalman estimator is determined given only the ratio of measurement noise to the process's underlying stochastic noise. Fixing both of these parameters allows Kalman methods to also identify the variance of the associated estimator. Furthermore, a line of research aims to identify parameters purely from data, including Barratt and Boyd [2020], van Breugel et al. [2020], Jonker et al. [2020]. Many methods include their own parameters and are not guaranteed a solution, but are an improvement on the indeterminate nature of direct maximum likelihood or MAP likelihood.

This paper introduces Kalman smoothing as the derivative estimation step in SINDy in distinction with the L1 total variation minimization or Savitzky-Golay smoothers common in application. It is not the first to combine Kalman methods with SINDy; Rosafalco et al. [2024] utilize Ensemble Kalman Filtering (EKF) to identify a partially-known system as a portion of a multi-step method, and Wang et al. [2022] apply Kalman filtering to the ODE coefficients as a way of modeling a non-stationary but separable system. This paper's introduction of Kalman smoothing a continuous process loss for derivative estimation, on the other hand, begins to align the derivative estimation step to the symbolic regression step. It allows engineering applications to incorporate SINDy estimation with a well-established and familiar data assimilation technique whose noise properties are well understood.

Section two describes the individual methods of SINDy and Kalman smoothing, providing some literature review. In section three, experiments demonstrate the advantages of incorporating Kalman with SINDy. The paper concludes with avenues for future research in section four.

II. BACKGROUND

A. SINDY

SINDy [Brunton et al., 2016] is a family of emerging methods for discovering the underlying dynamics of a system governed by unknown or partially-known [Champion et al., 2020] differential equations. It can handle ODEs as well as PDEs [Rudy et al., 2017], and has been used for protein folding [Boninsegna et al., 2018], chemical reaction networks [Hoffmann et al., 2019], plasma physics [Guan et al., 2021], and more. Most invocations occur through the pysindy Python package, but innovations such as Langevin Regression [Callaham et al., 2021] or Rudy et al. [2019] exist as independent code.

Given some variable of interest X and a library of functions Θ (including spatial derivatives, when relevant) SINDy seeks to find the coefficients Ξ of the differential equation:

$$\dot{X} = \Xi\Theta(X), \tag{1}$$

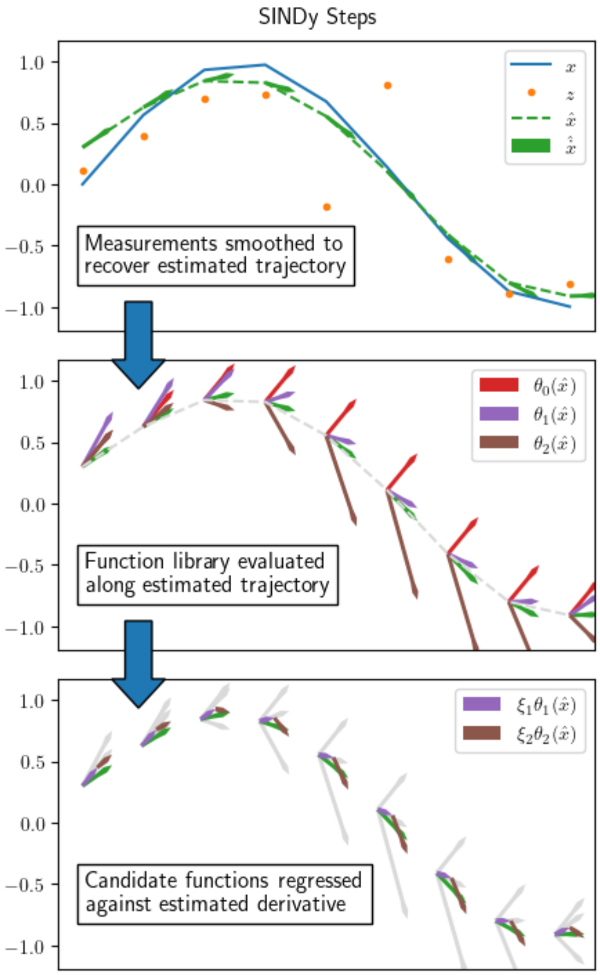

FIGURE 1. The SINDy method, applied to fitting a sinusoid. Taking noisy data, it identifies the model $\dot{x} = \xi_1\theta_1(x) + \xi_2\theta_2(x)$, rejecting θ_0.

where

$X \in \mathbb{R}^{n\times m} = x(t_1)...x(t_m)$: system of n coordinates at m timepoints

$\Theta(X) \in \mathbb{R}^{p\times m}$: library of p functions evaluated at m timepoints

$\Xi \in \mathbb{R}^{n\times p}$: coefficients for n equations of p functions

The function library written as a time-independent quantity refers to the collection $\Theta = [\theta_1, \ldots \theta_p]^T$, where $\theta_i : \mathbb{R}^n \to \mathbb{R}$. Examples include the family of all degree-2 polynomials of n inputs, mixed sines and cosines of certain frequencies, or any user-specified family.

The method generally presumes the measurements (Z) faithfully reflect system state (X) and proceeds in two steps:

1) Estimate the time derivatives of the system $\widehat{\dot{X}} = \mathrm{F}(Z)$ for some smoothing function F.
2) Choosing a sparse regression method, solve the problem $\underset{\text{sparse } \Xi}{\arg\min} \left\| \widehat{\dot{X}} - \Xi\Theta(X) \right\|^2$.

This general process is sketched out in Fig. 1. Researchers have tried a few different methods for calculating the derivatives, broadly grouped into global methods (e.g. L-1 total variation minimization of Chartrand et al. [2011]) and local methods (e.g. Savitzky-Golay smoothing). Different ways of applying sparsity has attracted more attention, including sequentially thresholding linear regression, nonconvex penalties such as L-0 with a relaxation-based minimization method [Champion et al., 2020, Zheng et al., 2018], an L-0 constraint [Bertsimas and Gurnee, 2023], and Bayesian methods for a prior distribution such as spike-slab or regularized horseshoe priors [Gao and Kutz, 2022, Hirsh et al., 2022]. The latter two papers also demonstrate an interesting line of innovation, eschewing derivatives and using the integral of function library in the loss term. A related approach instead uses the weak form of the differential equation, yielding a solution that *is* convex, but which does not provide as straightforward an interpretation of the measurement noise. Most of these methods can benefit from ensembling the data and library terms, as in Fasel et al. [2022], but others, such as Kaptanoglu et al. [2021] for identifying Galerkin modes of globally stable fluid flows, require a specific form of function library.

This paper seeks to make SINDy more resilient to noise by taking a data assimilation approach. It instead presents the Kalman SINDy steps:

1) Estimate the state and time derivatives of the system $\widehat{\dot{X}}, \widehat{X} = \mathrm{F}(X)$ where F applies Kalman smoothing.
2) Choosing a sparse regression method, solve the problem $\underset{\text{sparse } \Xi}{\arg\min} \left\| \widehat{\dot{X}} - \Xi\Theta(\widehat{X}) \right\|^2$.

B. KALMAN SMOOTHING

Kalman filtering and smoothing refers to a group of optimal estimation techniques to assimilate measurement noise to a random process. Filtering refers to incorporating new measurements in real-time, while smoothing refers to estimating the underlying state or signal using a complete trajectory of (batch) measurements. While the processes this paper is concerned with are not random, in the first step of SINDy they are unknown, and so probabilistic language is appropriate.

In adding Kalman smoothing to SINDy, we introduce a distinction between the measurement variables and the state variables of the dynamical system in equation 1. As such, the inputs to the problem become m time points of measurements of k variables ($Z \in \mathbb{R}^{k\times m}$) and a linear transform from the state to the measurement variables $H \in \mathbb{R}^{n\times k}$ describing how the process is measured.

Measurement error is assumed to be normally distributed with $HX - Z \sim \sigma_z \mathcal{N}(0, R)$ where the covariance matrix $R \in \mathbb{R}^{k\times k}$. Measurement regimes where noise is autocorrelated or varies over time can be accomodated by flattening $HX - Z$ and describing $R \in \mathbb{R}^{nk\times nk}$.

As a simplifying assumption for experiments in this paper, we use $R = I$. Two parameters are required: σ_z, the measurement noise standard deviation, and σ_x, the process velocity standard deviation per unit time. If only point estimates of the state are required, and posterior uncertainty is not, it suffices to use the ratio $\rho = (\sigma_z/\sigma_x)^2$.

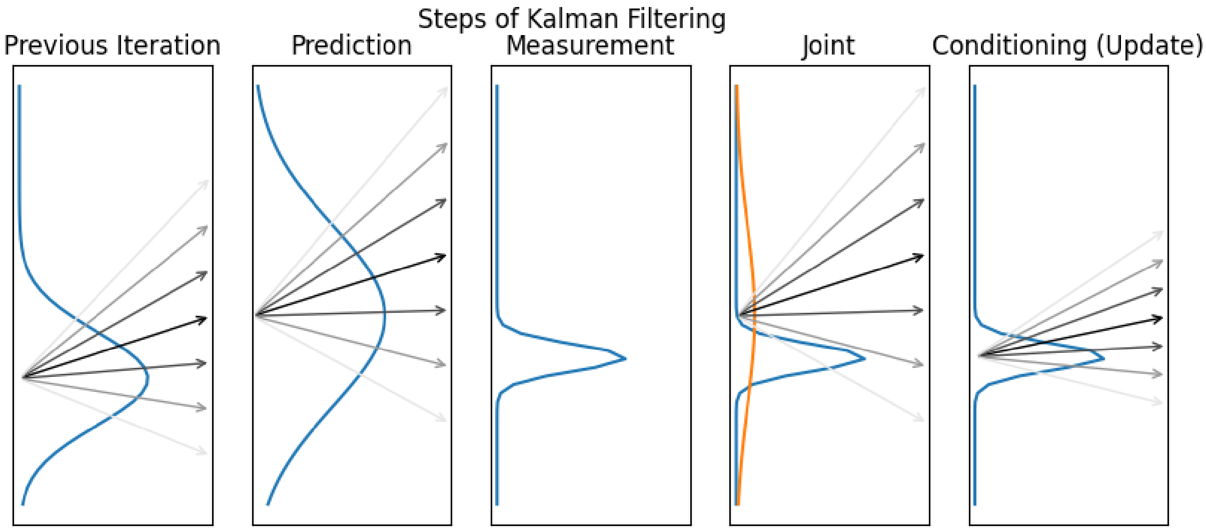

FIGURE 2. Explanatory depiction of Kalman filtering. A previous iteration gives a distribution $p(x_{i-1}, \dot{x}_{i-1})$. Multiplication by an update matrix produces the predictions $p(x_i, \dot{x}_i|x_{i-1}, \dot{x}_{i-1})$. Simultaneously, measurements z_i are taken that, with known measurement noise, give $p(z_i|x_i)$. Multiplication gives the joint distribution $p(x_i, \dot{x}_i, z_i|x_{i-1}, \dot{x}_{i-1})$, from which the conditional distribution $p(x_i, \dot{x}_i|z_i, x_{i-1}, \dot{x}_{i-1})$ can be calculated, shown in Eaton [2007]

Each process is assumed to have an independent, Brownian velocity. This leads to Kalman smoothing estimator:

$$\arg\min_{X,\dot{X}} \|HX - Z\|_{R^{-1}}{}^2 + \rho\|G[\dot{X}, X]\|_{Q^{-1}}{}^2. \tag{2}$$

Here, G is a linear transform to separate $[\dot{X}, X]$ into independent, mean-zero increments, and Q is the covariance of those increments. A graphic displaying Kalman filtering is shown in Fig. 2. To illustrate the ideas, the figure presents step-by-step filtering updates; however, batch smoothing is used for the model discovery applications presented in the experiments.

We use the generalized cross validation of Barratt and Boyd [2020] to choose ρ. This strategy chooses ρ in order to minimize the loss on a witheld set of data. While the algorithm described in that paper is not guaranteed to find a minimum, heuristic experience has shown that the longer the trajectory, the more likely their algorithm will succeed. The experiments in the next section show that this strategy works reasonably well.

The generalized cross validation approach of Barratt and Boyd [2020] witholds some measurement points in order to find the values of H, R, G, and Q that produce estimates $\hat{X}$ that fit witheld data most accurately. This powerful approach can apply to all linear systems but comes with the burdens of non-convexity. In our work, we presume to know the measurement parameters and most of the process parameters - after all, ‘‘position is the integral of velocity", implies certain constraints on G and Q. The method accomodates these constraints via specification of a prox function.

III. EXPERIMENTS

We seek to evaluate Kalman smoothing as a step in SINDy in comparison to other noise-mitigation innovations. We simulate eight dynamical systems[1] with noisy measurements across a variety of initial conditions, discovering ODEs from SINDy with different smoothing methods. The trials are run across a range of durations and relative noise levels, calculated as the noise-to-signal ratio of the measurement variance with the system's mean squared value. To compare the methods, we then integrate the discovered equations and observe how well they preserve the system's structure as well as directly comparing the coefficients through the F1 score and mean absolute error (MAE).

We compare the results of SINDy with Kalman smoothing and the hyperparameter optimization of Barratt and Boyd [2020] in comparison with alternative smoothing methods: L-1 total variation minimization and Savitzky-Golay. The latter smoothing methods have been modified to pass the not just the smoothed derivatives $\hat{\dot{X}}$, but also the smoothed position estimates $\hat{X}$ to the second step of SINDy. They also each require a parameter: TV requires a coefficient for the L-1 regularizer and Savitzky-Golay requires a smoothing window. These are gridsearched over a wide range, although it is worth noting that choosing the gridsearched optimum requires knowledge of the true system, in distinction to the hyperparameter optimization method used for Kalman smoothing. We also compare with a gridsearched Kalman smoothing to directly evaluate the efficacy of the generalized cross-validation hyperparameter selection.

Beyond the differentiation step, the SINDy models also

[1]Cubic Harmonic Oscillator, Duffing, Hopf, Lotka-Volterra, Rossler, Simple Harmonic Oscillator (SHO), Van Der Pol Oscillator, and Lorenz-63

specify a function library and optimizer. The feature library used for all experiments was cubic polynomials, including mixed terms and a constant term. The optimizer was the mixed-integer SINDy optimizer of Bertsimas and Gurnee [2023], configured with the correct number of nonzero terms a priori, and ensembled over 20 models each trained on 60% of the data. Presenting SINDy with the known number of nonzero coefficients is an attempt to present a best case, where we can ameliorate any interaction between the smoothing method and sparsification parameters. A full list of ODE, simulation, and experimental parameters are shown in the Appendix, tables I and II.

Methods can be compared in several ways: by the coefficients of the equations they discover, by their accuracy in forecasting derivatives, and how well the discovered system recreates observed dynamics in simulation. As Gilpin [2023] notes, there are many metrics for scoring dynamical system discovery, and the merit of a metric depends upon both the use case and whether the trajectory considered is one of importance. For instance, in controls engineering, the local derivative and very short-term forecasting is the primary imperative. On the other hand, for reduced-order PDE models, recreating larger-scale phenomena in simulation may be more important. Finally, in high-dimensional network dynamics, the accuracy of identifying connectivity, as measured by coefficient F1 score, is most important.

As the coefficient metrics are the most straightforwards, and we compare methods by F1 score and Mean Absolute Error as the duration of training data increases, and separately, as the measurement noise increases. We also visually evaluate how well the discovered ODEs, simulated from random initial conditions in a test set, track the true data and display relevant behavior.

A. RUNNING EXPERIMENTS

In a desire to make the experiments not just reproducible, but also reusable, we have separated the method, experiment, and experiment runner into separate packages. Methodological improvements include adding Kalman smoothing and a entry point for hyperparameter optimization to the `derivative` package, as well as an API for returning not just the derivatives, but the smoothed coordinates themselves (employed for Kalman and Total Variation). In `pysindy`, we enabled incorporating the smoothed coordinates into successive problem steps. It should be noted that previous experiments using `pysindy`'s derivative estimation would re-use the noisy coordinates in function library evaluation.

Within `pysindy`, we redefined ensembling in a more flexible way to apply to a greater variety of underlying optimizers, such as the MIOSR one used in these experiments. The standardization of interfaces allows us to compose SINDy experiments in the `pysindy-experiments` package [Stevens-Haas and Bhangale, 2024]. It allows a standard API to specify data generation, model building, and evaluation.

Finally, in order to make it easier to collaborate and reproduce experiments, we expanded the `mitosis` package [Stevens-Haas, 2024]. This package allows specifying experiment parameters and groups in a declarative manner, which leads to more readable diffs. It also pins reproducibility information for any experiment run. Further reproducibility info is in the Appendix.

B. RESULTS

We find that SINDy with Kalman smoothing recovers the problem structure in application as well or better than competing methods. Models discovered in this manner track the essential dynamics in most cases. SINDy with Kalman hyperparameter optimization tends to perform worse than that with Savitzky-Golay, but on par with Total Variation gridsearched optima, and is itself outperformed only slightly by the Kalman gridsearched optima. While hyperparameter optimization imposes some runtime cost, it does not require access to the true data, making those results all the more inspiring for field use cases. Simulations of discovered models across all ODEs and methods are shown in Fig. 3

Surprisingly, methods that smooth better, as shown in Fig. 4 do not necessarily recreate the essential dynamics better in simulation. As a case in point, the Kalman-smoothed training data itself does not seem as accurate as data smoothed by L-1 Total Variation in the the SHO trial.

Even more surprisingly, despite performing well in simulation, SINDy with Kalman hyperparameter optimization performs middlingly in the coefficient metrics. If there's anything that appears consistent about Kalman with GCV, it is that, with a long enough duration, performance appears insensitive regardless of noise, as shown in Fig. 6. Across the range of noise levels sampled, either Savitzky-Golay or Kalman (gridsearched) perform the best, depending upon system. As expected, Kalman (gridsearched) always outperforms Kalman GCV, but it is interesting to note that at some durations and noise levels Kalman GCV occasionally outperforms Savitzky-Golay (e.g. Rossler, Hopf). Coefficient metrics by data duration is shown in Fig. 5.

Generally, MAE seems to provide a better indication of which method will perform better in simulation than F1 score. Nevertheless, there are cases where the MAE scores of different methods do not indicate which method performs better in simulation, and where effective smoothing does not predict effective system recovery. As one case in point, Kalman GCV and Total Variation smoothing appears most accurate for the Hopf system in Fig. 4. However, the coefficient metrics show that either Kalman or Savitzky-Golay recovered the system equations better, and Fig. 3 shows that Savitzky-Golay reconstructed the dynamics more accurately. Similarly, methods have a wide range of performance on MAE and F1 score on the Rossler system, despite all simulations missing the chaotic behavior.

IV. CONCLUSION

This paper has demonstrated that Kalman smoothing is a useful addition to SINDy. It makes the method more generally applicable across domains. The Kalman smoother behaves

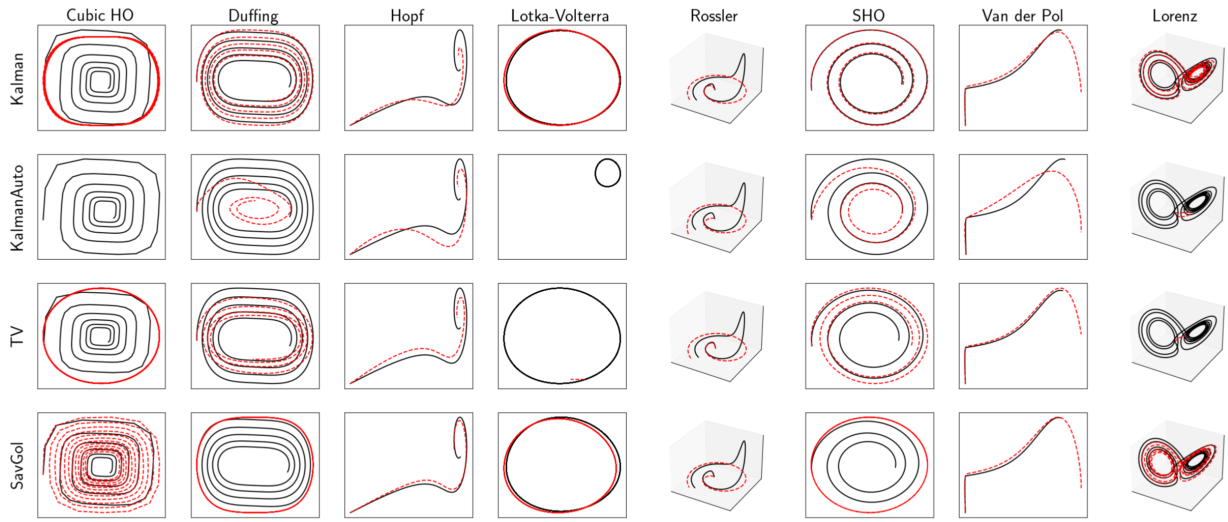

FIGURE 3. The simulation of discovered models compared to test data. Kalman appears better for half of eight ODEs. It represents the essential behavior of more ODEs than TV and Savitzky-Golay. Kalman with auto-hyperparameter selection performs similarly to total variation on a gridsearch. 10% relative noise, 8 seconds of data.

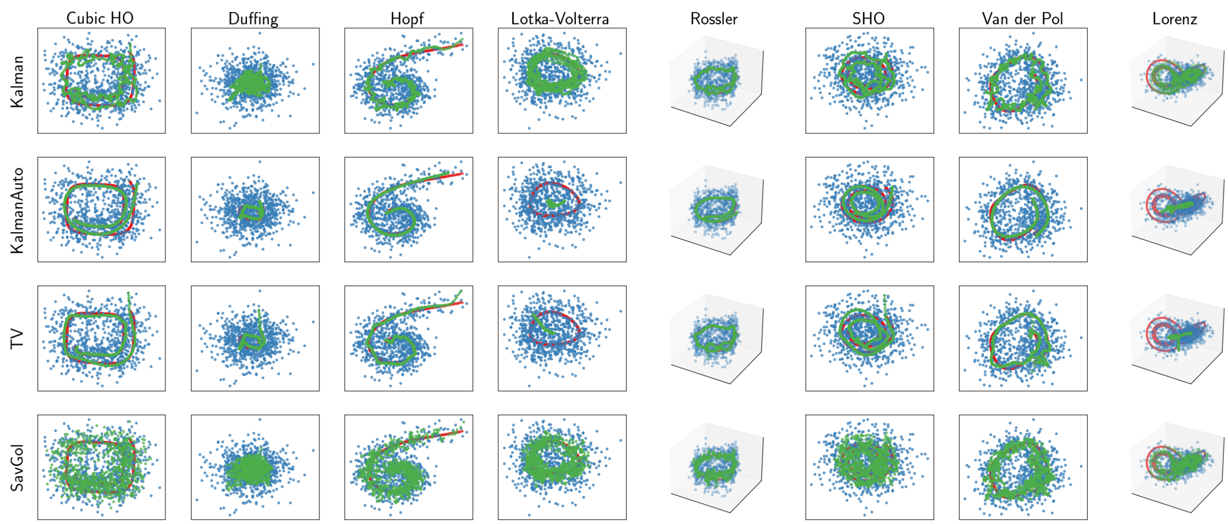

FIGURE 4. The smoothing of training data, performed by different differentiation methods prior to SINDy fit. It does not appear to be the case that a more visually accurate smoothing yields a model that behaves more correctly in simulation. Nevertheless, as Fig. 3 shows, Kalman-smoothed trajectories lead to better models in simulation. 10% relative noise, 8 seconds of data.

optimally for the simplest systems and provides a familiar process to the controls engineering community. It also appears to perform better at preserving global system structure in simulation. Incorporating the GCV hyperparameter optimization of Barratt and Boyd [2020] may not recover the best model, but it allows one to at least recover useful models without relying on an accurate parametrization a priori, particularly if substantial training data exists. However, ‘‘best model" means different things for different use cases. For uncovering connections between variables, such as in neural activity or chemical reaction networks Hoffmann et al. [2019], the performance on coefficient F1 metrics indicates that more accurate parameter tuning is essential.

The field is rife with a diversity of follow up studies. Firstly, since Kalman smoothing and SINDy regression loss terms both accommodate variable timesteps, a natural innovation is to combine the two into a single optimization problem. Hirsh et al. [2022] and Rudy et al. [2019] introduce a single-step optimization, but do not evaluate their single step methods in comparison to the mathematically nearest two-step SINDy. As a result, it is difficult to evaluate that aspect of their innovations in isolation. Following this line of inquiry, producing a single-step SINDy that utilizes Kalman loss could allow a more clear trade-off between measurement noise or the coefficient sparsity.

In parallel, more could be done to give hyper-parameter optimization access to the terms in the SINDy expression. The method of Barratt and Boyd [2020] is supremely general,

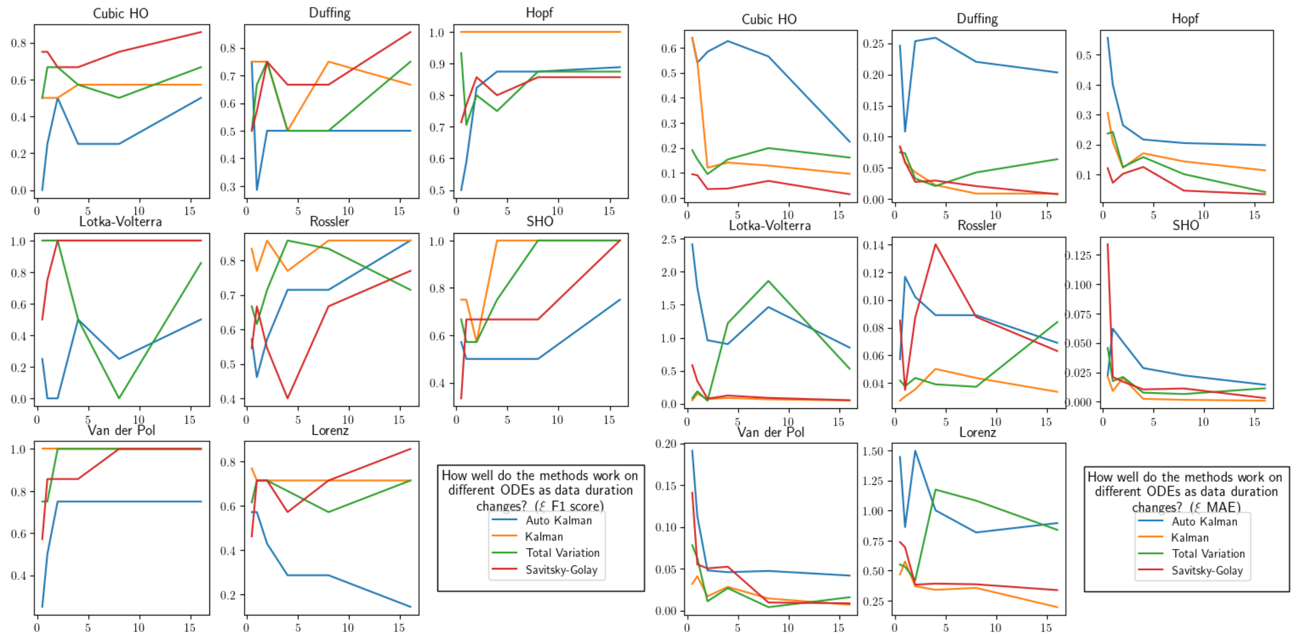

FIGURE 5. How well different smoothing methods in SINDy recover the ODE coefficients as data duration increases

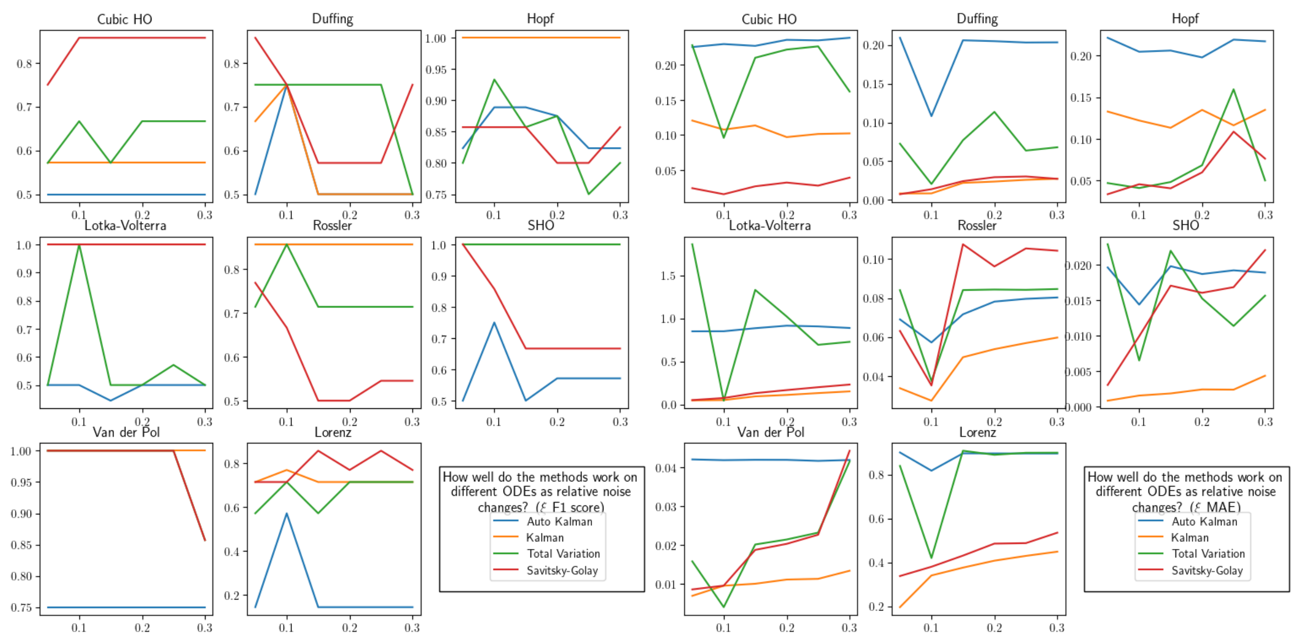

FIGURE 6. How well different smoothing methods in SINDy recover the ODE coefficients as noise increases

with no intrinsic understanding of the process variance or measurement noise. Applying that knowledge to their prox-gradient method was part of this paper. However, the method is nonconvex, which became problematic with the restriction to the scalar ρ in equation 2. Moreover, on the path to a single-step SINDy lies an opportunity to use knowledge of the ODE terms in hyper-parameter optimization. In a related note, van Breugel et al. [2020] also provide hyper-parameter estimation techniques for Savitzky-Golay that could be evaluated in the experiments of this paper.

Kalman SINDy could also be more directly compared to Weak SINDy [Messenger and Bortz, 2021a], which aims at the same goal of reducing the sensitivity to noise. However, it's implementation in pysindy does not allow for simulation, and modifying existing code to provide that comparison is an investigation in its own right and a necessary next step.

Finally, the interpretation of Kalman smoothing as the maximum likelihood estimator for Brownian motion suggests that it could inform the attempts at a stochastic SINDy, for which Callaham et al. [2021] and Boninsegna et al. [2018] have made the first steps. Stochastic SINDy in those cases aimed for the use cases of noise arising from PDE discretization and inherently statistical mechanics, but the general formulation also has use in any randomly-forced system, such as HVAC controls for a building or mapping a limited part of a chemical reaction network.

APPENDIX.

This paper is built from https://github.com/Jacob-Stevens-Haas/Kalman-SINDy-paper. To run the experiments, install the package located in images/gen_image and run the commands in images/gen_image/run_exps.sh. Each experiment trial will generate a pseudorandom hex key for reproducibility. To build the final figures, edit images/gen_image/composite_plots.py with the keys to the experimental results and run it.

While the exact parametrization is in the experimental configuration and package defaults, it is recreated here in Tables I and II.

REFERENCES

Steven L Brunton, Joshua L Proctor, and J Nathan Kutz. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proceedings of the National Academy of Sciences*, 113, 2016. .

Steven L Brunton and J Nathan Kutz. *Data-driven science and engineering: Machine learning, dynamical systems, and control*. Cambridge University Press, 2022.

Lorenzo Boninsegna, Feliks Nüske, and Cecilia Clementi. Sparse learning of stochastic dynamical equations. *Journal of Chemical Physics*, 148, 6 2018. ISSN 00219606. .

Samuel H Rudy, Steven L Brunton, Joshua L Proctor, and J Nathan Kutz. Data-driven discovery of partial differential equations. *Science Advances*, 3, 2017. URL https://www.science.org.

R E Kalman. A New Approach to Linear Filtering and Prediction Problems. *Transactions of the AMSE - Journal of Basic Engineering*, 82(D):35–45, 1960.

R E Kalman and R S Bucy. New results in linear filtering and prediction theory. *Trans. ASME J. Basic Eng*, 83:95–108, 1961.

Greg Welch, Gary Bishop, et al. *An introduction to the Kalman filter*. Chapel Hill, NC, USA, 1995.

J Nathan Kutz, Steven L Brunton, Bingni W Brunton, and Joshua L Proctor. *Dynamic mode decomposition: data-driven modeling of complex systems*. SIAM, 2016.

Sara M. Ichinaga, Francesco Andreuzzi, Nicola Demo, Marco Tezzele, Karl Lapo, Gianluigi Rozza, Steven L. Brunton, and J. Nathan Kutz. Pydmd: A python package for robust dynamic mode decomposition, 2024.

M Raissi, P Perdikaris, and G E Karniadakis. 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, 2019. . URL www.elsevier.com/locate/jcp.

Salvatore Cuomo, Vincenzo Schiano Di Cola, Fabio Giampaolo, Gianluigi Rozza, Maziar Raissi, and Francesco Piccialli. Scientific machine learning through physics–informed neural networks: Where we are and what's next. *Journal of Scientific Computing*, 92(3):88, 2022.

Joshua S North, Christopher K Wikle, and Erin M Schliep. A review of data-driven discovery for dynamic systems. *International Statistical Review*, 91(3):464–492, 2023.

Daniel A. Messenger and David M. Bortz. Weak sindy for partial differential equations. *Journal of Computational Physics*, 443, 10 2021a. ISSN 10902716. .

Daniel A. Messenger and David M. Bortz. Weak sindy: Galerkin-based data-driven model selection. *Multiscale Modeling and Simulation*, 19:1474–1497, 2021b. ISSN 15403467. . URL https://doi.org/10.1137/20M1343166.

Hayden Schaeffer and Scott G Mccalla. Sparse model selection via integral terms. *PHYSICAL REVIEW E*, 96:23302, 2017. .

U. Fasel, J. N. Kutz, B. W. Brunton, and S. L. Brunton. Ensemble-sindy: Robust sparse model discovery in the low-data, high-noise limit, with active learning and control. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 478, 2022. ISSN 14712946. .

L Gao and J Nathan Kutz. Bayesian autoencoders for data-driven discovery of coordinates, governing equations and fundamental constants. *arXiv preprint arXiv:2211.10575*, 2022.

L Gao, Urban Fasel, Steven L Brunton, and J Nathan Kutz. Convergence of uncertainty estimates in ensemble and bayesian sparse model discovery. *arXiv preprint arXiv:2301.12649*, 2023.

Seth M. Hirsh, David A. Barajas-Solano, and J. Nathan Kutz. Sparsifying priors for bayesian uncertainty quantification in model discovery. *Royal Society Open Science*, 9, 2022. ISSN 20545703. . URL https://doi.org/10.1098/rsos.211823.

Alan Kaptanoglu, Brian de Silva, Urban Fasel, Kadierdan Kaheman, Andy Goldschmidt, Jared Callaham, Charles Delahunt, Zachary Nicolaou, Kathleen Champion, Jean-Christophe Loiseau, J. Kutz, and Steven Brunton. Pysindy: A comprehensive python package for robust sparse system identification. *Journal of Open Source Software*, 7:3994, 1 2022. . A summary of variants to SINDy as of late 2021. Addition of PDEs, ensembling, and a few optimizers (Trapping, FROLS, PI).

Zichao Long, Yiping Lu, and Bin Dong. Pde-net 2.0: Learning pdes from data with a numeric-symbolic hybrid deep network. *Journal of Computational Physics*, 399:108925, 2019. . URL www.elsevier.com/locate/jcp.

Zichao Long, Yiping Lu, Xianzhong Ma, and Bin Dong. Pde-net: Learning pdes from data. In *Proceedings of the International Conference on Machine Learning*, 2018.

Steven Atkinson. Bayesian hidden physics models: Un-

System	ODE	Experiment parameters	x_0 mean
Linear Damped Oscillator	$\dot{x} = \begin{bmatrix} -\alpha & \beta \\ -\beta & -\alpha \end{bmatrix} x$	(0,0)	$\alpha = .1, \beta = 2$
Lorenz	$\dot{x} = \begin{bmatrix} \sigma(x_2 - x_1) \\ x_1(\rho - x_3) - x_2 \\ x_1 x_2 - \beta x_3 \end{bmatrix}$	$\sigma = 10, \rho = 28, \beta = 8/3$	(0, 0, 15)
Cubic Damped Oscillator	$\dot{x} = \begin{bmatrix} -\alpha & \beta \\ -\beta & -\alpha \end{bmatrix} \begin{bmatrix} x_1^3 \\ x_2^3 \end{bmatrix}$	$\alpha = .1, \beta = 2$	(0,0)
Duffing	$\dot{x} = \begin{bmatrix} x_2 \\ -\alpha x_2 - \beta x_1 - \gamma x_1^3 \end{bmatrix}$	$\alpha = .2, \beta = .05, \gamma = 1$	(0,0)
Hopf	$\dot{x} = \begin{bmatrix} -\alpha x_1 - \beta x_2 - \gamma x_1(x_1^2 + x_2^2) \\ \beta x_1 - \alpha x_2 - \gamma x_2(x_1^2 + x_2^2) \end{bmatrix}$	$\alpha = .05, \beta = 1, \gamma = 1$	(0,0)
Lotka-Volterra	$\dot{x} = \begin{bmatrix} \alpha x_1 - \beta x_1 x_2 \\ \beta x_1 x_2 - 2\alpha x_2 \end{bmatrix}$	$\alpha = 5, \beta = 1$	$(5, 5)$
Rossler	$\dot{x} = \begin{bmatrix} -x_2 - x_3 \\ x_1 + \alpha x_2 \\ \beta + (x_1 - \gamma) x_3 \end{bmatrix}$	$\alpha = .2, \beta = .2, \gamma = 5.7$	(0,0,0)
Van der Pol Oscillator	$\dot{x} = \begin{bmatrix} x_2 \\ \alpha(1 - x_1^2) x_2 - x_1 \end{bmatrix}$	$\alpha = .5$	(0,0)

TABLE I. The parametrization of ODEs used in these experiments. Mostly from defaults in the pysindy package.

Parameter	Value
Simulated Data	
Number of trajectories	10
Initial Condition (x_0) variance	9
Initial Condition (x_0) distribution	Normal*
Measurement error elative noise (default)	10%
Trajectory duration (default)	16
Measurement interval	0.01
Random seed	19
SINDy model	
Feature Library (Θ)	Polynomials to degree 3
Optimizer	Mixed Integer Optimizer
L2 regularization (coefficitents) (α)	0.01
Target sparsity	(true value from equation)
Unbiasing	Yes
Feature normalization	No
Ensembling	data bagging
Number of bags	20
Experiment	
Trajectory duration (grid)	0.5, 1, 2, 4, 8, 16
Relative noise (grid)	0.05, 0.1, 0.15, 0.2, 0.25, 0.3
Measurement:Process variance (Kalman grid)	1e-4, 1e-3, 1e-2, 1e-1, 1
L1 regularization (derivative) (TV grid)	1e-4, 1e-3, 1e-2, 1e-1, 1
Window length (Savitzky-Golay grid)	5, 8, 12, 15

TABLE II. Parametrization of data, SINDy models, and experiments conducted.
***Lotka-Volterra uses a gamma distribution, rather than normal, in order to enforce nonnegativity.**

certainty quantification for discovery of nonlinear partial differential operators from data, 2020. URL https://arxiv.org/abs/2006.04228.

Miles D Cranmer, Rui Xu, Peter Battaglia, and Shirley Ho. Learning symbolic physics with graph networks. *33rd annual conference on Neural Information Processing Systems*, 2019. URL https://ml4physicalsciences.github.io/2019/files/NeurIPS_ML4PS_2019_15.pdf.

Miles Cranmer, Alvaro Sanchez-Gonzalez, Peter Battaglia, Rui Xu, Kyle Cranmer, David Spergel, and Shirley Ho. Discovering symbolic models from deep learning with inductive biases. *Advances in Neural Information Processing Systems 33*, 2020. URL https://github.com/MilesCranmer/symbolic_.

Alvaro Sanchez-Gonzalez, Jonathan Godwin, Tobias Pfaff, Rex Ying, Jure Leskovec, and Peter Battaglia. Learning to simulate complex physics with graph networks. In Hal Daumé III and Aarti Singh, editors, *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pages 8459–8468. PMLR, 13–18 Jul 2020. URL https://proceedings.mlr.press/v119/sanchez-gonzalez20a.html.

Aleksandr Aravkin, James V Burke, Lennart Ljung, Aurelie Lozano, and Gianluigi Pillonetto. Generalized kalman smoothing: Modeling and algorithms. *Automatica*, 86:63–86, 2017.

Aleksandr Aravkin, James V Burke, and Gianluigi Pillonetto. Robust and trend-following kalman smoothers using student's t. *IFAC Proceedings Volumes*, 45(16):1215–1220, 2012.

Jonathan Jonker, Aleksandr Aravkin, James V Burke, Gianluigi Pillonetto, and Sarah Webster. Fast robust methods for singular state-space models. *Automatica*, 105:399–405, 2019.

Shane T. Barratt and Stephen P. Boyd. Fitting a kalman smoother to data. In *2020 American Control Conference (ACC)*, pages 1526–1531, 2020. .

F. van Breugel, J. Nathan Kutz, and B. W. Brunton. Numerical differentiation of noisy data: A unifying multi-objective optimization framework. *IEEE Access*, 2020. .

Jonathan Jonker, Peng Zheng, and Aleksandr Y Aravkin. Efficient robust parameter identification in generalized kalman smoothing models. *IEEE Transactions on Automatic Control*, 66(10):4852–4857, 2020.

Luca Rosafalco, Paolo Conti, Andrea Manzoni, Stefano Mariani, and Attilio Frangi. EKF-SINDy: Empowering the extended kalman filter with sparse identification of nonlinear dynamics, 2024. URL http://arxiv.org/abs/2404.07536.

Jingyi Wang, Jesus Moreira, Yankai Cao, and Bhushan Gopaluni. Time-Variant Digital Twin Modeling through the Kalman-Generalized Sparse Identification of Nonlinear Dynamics. In *2022 American Control Conference (ACC)*, pages 5217–5222, Atlanta, GA, USA, June 2022. IEEE. ISBN 978-1-66545-196-3. . URL https://ieeexplore.ieee.org/document/9867786/.

Kathleen Champion, Peng Zheng, Aleksandr Y. Aravkin, Steven L. Brunton, and J. Nathan Kutz. A unified sparse optimization framework to learn parsimonious physics-informed models from data. *IEEE Access*, 8:169259–169271, 2020. ISSN 21693536. .

Moritz Hoffmann, Christoph Fröhner, and Frank Noé. Reactive sindy: Discovering governing reactions from concentration data. *Journal of Chemical Physics*, 150, 1 2019. ISSN 00219606. .

Yifei Guan, Steven L. Brunton, and Igor Novosselov. Sparse nonlinear models of chaotic electroconvection. *Royal Society Open Science*, 8, 8 2021. ISSN 20545703. .

J. L. Callaham, J. C. Loiseau, G. Rigas, and S. L. Brunton. Nonlinear stochastic modelling with langevin regression. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 477, 6 2021. ISSN 14712946. .

Samuel H Rudy, Steven L Brunton, and J Nathan Kutz. Smoothing and parameter estimation by soft-adherence to governing equations. *Journal of Computational Physics*, 398:108860, 2019. . URL www.elsevier.com/locate/jcp.

Rick Chartrand, L Marin, and D Xiao. Numerical differentiation of noisy, nonsmooth data. *International Scholarly Research Network ISRN Applied Mathematics*, 2011:11, 2011. .

Peng Zheng, Travis Askham, Steven L Brunton, J Nathan Kutz, and Aleksandr Y Aravkin. A unified framework for sparse relaxed regularized regression: Sr3. *IEEE Access*, 7:1404–1423, 2018.

Dimitris Bertsimas and Wes Gurnee. Learning sparse nonlinear dynamics via mixed-integer optimization. *Nonlinear Dynamics*, 111:6585–6604, 4 2023. ISSN 1573269X. .

Alan A Kaptanoglu, Jared L Callaham, Aleksandr Aravkin, Christopher J Hansen, and Steven L Brunton. Promoting global stability in data-driven models of quadratic nonlinear dynamics. *PHYSICAL REVIEW FLUIDS*, 6:94401, 2021. .

Morris L. Eaton. *Multivariate Statistics: A Vector Space Approach*, volume 53. Institute of Mathematical Statistics, 2007. ISBN 9780940600690. . URL https://projecteuclid.org/ebooks/institute-of-mathematical-statistics-lecture-notes-monograph-series Multivariate-Statistics/toc/10.1214/lnms/1196285102.

William Gilpin. Chaos as an interpretable benchmark for forecasting and data-driven modelling, 2023. URL https://github.com/williamgilpin/dysts.

Jacob Stevens-Haas and Yash Bhangale. pysindy-experiments, April 2024. URL https://github.com/Jacob-Stevens-Haas/gen-experiments.

Jacob Stevens-Haas. mitosis, April 2024. URL https://github.com/Jacob-Stevens-Haas/mitosis.

JAKE M. STEVENS-HAAS received the B.S. degree in Quantitative Economics from the United States Naval Academy in 2011.

From 2011 to 2017, he was Auxiliaries officer on the USS Momsen (DDG-92) and Weapons and Tactics instructor for Destroyer Squadron NINE. He is currently a PhD. Candidate in Applied Mathematics at University of Washington.

Mr. Stevens-Haas is the lead maintainer of `pysindy`, the python package for Sparse Identification of Nonlinear Dynamics, as well as various patches to various data science Python packages.

YASH BHANGALE , born in Jalgaon, Maharashtra, India on May 6th 1999. He received B. Tech. degree in mechanical engineering from the MIT World Peace University at Pune, Maharashtra, India in 2021, is currently pursuing M.S. in mechanical engineering at the University of Washington, Seattle, Washington, United States. His research interests include Dynamical Systems, Machine Learning, System Identification and Differential Equations.

J. NATHAN KUTZ (Senior Member, IEEE) received the B.S. degree in physics and mathematics from the University of Washington, Seattle, WA, USA, in 1990, and the Ph.D. degree in applied mathe- matics from Northwestern University, Evanston, IL, USA, in 1994. He is currently Director of the AI Institute in Dynamics Systems, a Professor of applied mathematics and electrical engineering, and a Senior Data Science Fellow with the eScience Institute, University of Washington.

ALEKSANDR ARAVKIN received B.S. degrees in Mathematics and Computer Science in 2004, an M.S. in Statistics and a Ph.D. in Mathematics in 2010, from the University of Washington. In 2015, Dr. Aravkin rejoined UW, where he is currently Associate Professor of Applied Mathematics, Adjunct Associate Professor in Statistics, Computer Science, Mathematics, and Health Metrics sciences, Data Science Fellow at the eScience Institute and Director of Mathematical Sciences at the Institute for Health Metrics and Evaluation