Letter   Featured in Physics

Data-driven discovery and extrapolation of parameterized pattern-forming dynamics

Zachary G. Nicolaou ,[1] Guanyu Huo ,[1] Yihui Chen ,[1] Steven L. Brunton ,[2] and J. Nathan Kutz[1]

[1]*Department of Applied Mathematics, University of Washington, Seattle, Washington 98195, USA*
[2]*Department of Mechanical Engineering, University of Washington, Seattle, Washington 98195, USA*

Pattern-forming systems can exhibit a diverse array of complex behaviors as external parameters are varied, enabling a variety of useful functions in biological and engineered systems. First-principles derivations of the underlying transitions can be characterized using bifurcation theory on model systems whose governing equations are known. In contrast, data-driven methods for more complicated and realistic systems whose governing evolution dynamics are unknown have only recently been developed. Here, we develop a data-driven approach, the *sparse identification for nonlinear dynamics with control parameters* (SINDyCP), to discover dynamics for systems with adjustable control parameters, such as an external driving strength. We demonstrate the method on systems of varying complexity, ranging from discrete maps to systems of partial differential equations. To mitigate the impact of measurement noise, we also develop a weak formulation of SINDyCP and assess its performance on noisy data. We demonstrate applications including the discovery of universal pattern-formation equations, and their bifurcation dependencies, directly from data accessible from experiments and the extrapolation of predictions beyond the weakly nonlinear regime near the onset of an instability.

Data-driven approaches to system identification are undergoing a revolution, spurred by the increasing availability of computational resources, data, and the development of novel and reliable machine learning algorithms [1–3]. The *sparse identification of nonlinear dynamics* (SINDy) is a particularly simple and flexible mathematical approach that leverages efficient sparse optimization algorithms in the automated discovery of complex system dynamics and governing equations [4]. In this Letter, we leverage the SINDy model discovery framework to understand parametric dependencies and underlying bifurcations in pattern-forming systems. Specifically, we develop SINDY with control parameters (SINDyCP) to discover such parameterized dynamics.

It has been 30 years since Cross and Hohenberg's seminal and authoritative review consolidating an exceptionally large body of work on pattern formation across a broad range of physical systems [5]. Universal equations determined by normal forms of canonical bifurcations [6], such as the complex Ginzburg-Landau equation [7], govern the formation of patterns near the onset of instabilities across scientific disciplines. Such equations continue to reveal insights into complex systems, including in the study of, for example, synchronization, biophysics, active matter, and quantum dynamics [8,9].

Despite the success of pattern-formation theory in modeling complex dynamics, ongoing challenges remain in applying such model equations more broadly. First-principles derivations and the computation of normal-form parameters in terms of physical driving parameters are tedious, costly, and error prone. Furthermore, the resulting weakly nonlinear models are only theoretically justified near the onset of instability, while interesting and important pattern-forming processes often occur far from the instability threshold. Recent advances in data-driven system identification are opening new avenues of research to address these challenges, including a paradigm for modeling strongly nonlinear regimes beyond the asymptotic approximations reviewed by Cross and Hohenberg [5].

The SINDy model discovery framework is particularly well suited to the modern analysis of bifurcations and normal forms, as it generates interpretable models that have as few terms as possible, balancing model complexity and descriptive capability. A variety of extensions of the SINDy approach have been developed since its introduction. For example, SINDYc enables the discovery of systems subject to external control signals [10–12], while PDEFind [13,14] enables the discovery of spatiotemporal dynamics characterized by partial differential equations (PDEs). Data-driven approaches can also learn to disambiguate between parametric dependency and governing equations and discover bifurcations [4,15,16]. Model pattern-formation equations typically encode the effects of external drive through a number of driving parameters, which characterize the bifurcation leading to the onset of instability. Several recent works establish system identification on pattern-forming systems ranging from closure models for fluid turbulence [17–20] to biochemical reactions and active matter systems [21–23]. These approaches show promise, but *crucially*, they have not demonstrated the ability to extrapolate by detecting pattern-forming instabilities that may develop when driving parameters differ from those used in the training data.
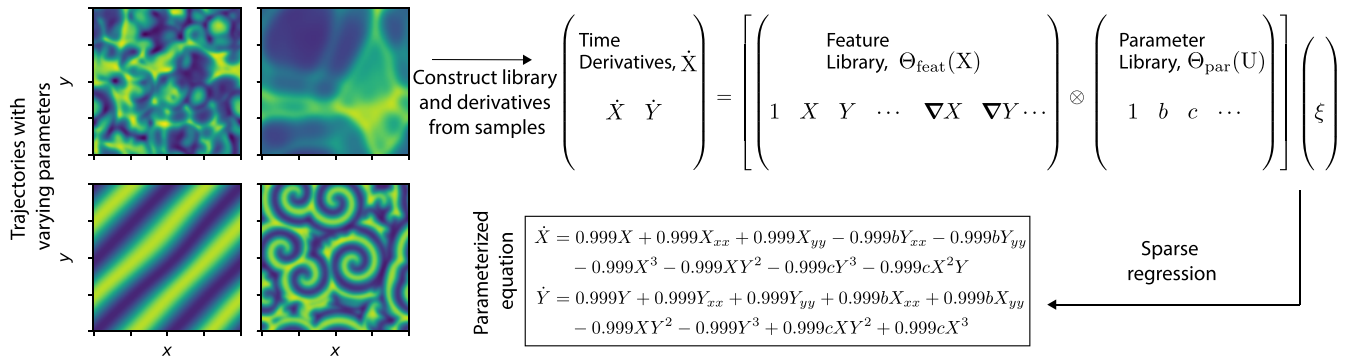
FIG. 1. Schematic of the SINDyCP approach. Data collected from sample trajectories collected under various driving parameters are processed to construct a matrix of time derivatives, a feature library Θ_{feat} of possible governing terms, and a parameter library Θ_{par} of parametric dependencies. Sparse regression is applied to the library coefficients ξ to identify a parameterized governing equation.

Our approach is implemented in the open-source PySINDy repository [24,25], enabling other powerful methods to be used in conjunction (see Supplemental Material Sec. S1A [30]). In particular, we develop and assess a weak formulation [26–29] of SINDyCP, which shows excellent performance on noisy data. We demonstrate that the method can be easily and effectively employed to discover accurate parameterized models from the kind of data available in typical pattern-formation experiments and that these parameterized models enable extrapolation beyond the conditions under which they were developed.

Building the library. Figure 1 illustrates the SINDyCP approach applied to the spatiotemporal evolution of four trajectories of the complex Ginzburg-Landau equation

$$\dot{A} = A + (1 + ib)\nabla^2 A - (1 - ic)|A|^2 A, \qquad (1)$$

which is described by a complex dependent variable $A(\mathbf{x}, t)$ in two spatial dimensions $\mathbf{x} = (x, y)$. Ginzburg-Landau exhibits a stunning variety of patterns, depending on the bifurcation parameters b and c. We generate four trajectories with parameter values $(b, c) = (2.0, 1.0)$, $(2.0, 0.75)$, $(0.5, 0.5)$, and $(1.0, 0.75)$, which exhibit differing dynamical phases, corresponding to amplitude turbulence, phase turbulence, stable waves, and frozen spiral glasses, respectively [7]. Our goal is to discover the partial differential equation for the real and imaginary components $A = X + iY$ parameterized by b and c from time series data.

As with most SINDy algorithms, we first form a matrix of the input data X, whose columns correspond to the dependent variables and whose rows correspond to the sample measurements of the dependent variables. In the case of Fig. 1, for example, X consists of two columns corresponding to the real and imaginary parts of A and $4N_x N_y N_t$ rows, where N_x, N_y, and N_t are the number of sample points in the corresponding spatiotemporal dimensions; again, there are four trajectories. We then determine the temporal derivative $\dot{X}$ for each sample point, either through numerical differentiation or through direct measurements.

In basic SINDy, we define a matrix of library terms $\Theta = \Theta(X)$ depending on the input data, which includes all possible terms that may be present in the differential equation that describes the temporal derivatives. These terms may be built from polynomial combinations of the dependent variables and

their spatial derivatives, for example, although more general libraries are possible. In the SINDYc approach, we augment the library dependence with an external control signal U, i.e., $\Theta = \Theta(X, U)$. The library terms are typically determined by appending the control variables to the dependent variables and again forming polynomials and derivatives. In the case in Fig. 1, we can treat the parameters as external control signals, $U = (b, c)$ and apply SINDYc, but the traditional implementation of this approach will fail for PDEs, as we show.

SINDYc aims to find a sparse linear combination of the library terms determined by the vector of coefficients ξ which minimizes the fit error

$$\xi^* = \underset{\xi}{\mathrm{argmin}} \, |\dot{X} - \Theta(X, U)\xi| + \lambda |\xi|_0, \qquad (2)$$

where sparsity promoting regularizing term $\lambda |\xi|_0$ penalizes nonzero coefficients via the $L0$ norm. Crucially, all SINDy methods employ sparse regression (with appropriate regularization) to determine a sparse set of nonzero coefficients ξ^*. Such sparsity is expected in physically relevant dynamics and produces parsimonious and interpretable models. Here, we employ the sequentially thresholded least-squares algorithm [4], which iteratively eliminates library terms with coefficients that fall below a threshold hyperparameter.

One challenge that arises when applying the traditional SINDYc to control parameters in PDEs with existing implementations such as PySINDy is that the matrix of library terms Θ is traditionally formed by computing *all* polynomial combinations of spatial derivatives of the dependent and control variables. However, since the control parameters are spatially constant, the spatial derivatives will vanish identically, leading to a singular matrix Θ. Such degeneracies lead to poor numerical results on data with control parameters when combining PDEFind and SINDYc without modification. To overcome this challenge, we propose constructing a more general library through products of a feature library $\Theta_{\text{feat}}(X)$ and a parameter library $\Theta_{\text{par}}(U)$, as

$$\Theta(X, U) = \Theta_{\text{feat}}(X) \otimes \Theta_{\text{par}}(U), \qquad (3)$$

where the product $\otimes$ here is defined to give the matrix consisting of all combinations of products of columns between the libraries, i.e., the ith row of $A \otimes B$ contains all the products of the form $A_{ij} B_{ik}$ where j and k span the columns of A and

B, respectively. By distinguishing the feature and parameter library dependencies with this SINDyCP approach, we can construct much more targeted and well-conditioned libraries.

Using a feature library consisting of spatial derivatives up to third order and polynomials up to third order along with a linear parameter library, the SINDyCP approach easily discovers Eq. (1) in Cartesian coordinates, as shown in Fig. 1. Additional demonstrations of SINDyCP for maps and ordinary differential equations (ODEs) are available in the Supplemental Material Sec. S1B, and further details of the complex Ginzburg-Landau equation (CGLE) integration, along with an animation illustrating the temporal evolution of the sample trajectories, are available in the Supplemental Material Sec. S2 [30].

Amplitude dynamics beyond weakly nonlinear theory. To illustrate the application of SINDyCP to pattern formation, we implement a numerical demonstration with the Belousov-Zhabotinksy chemical reaction system. We numerically integrate the Oregonator model [31],

$$\dot{C}_X = k_1 C_A C_H^2 C_Y - k_2 C_H C_X C_Y + k_3 C_A C_H C_X$$
$$\quad - 2k_4 C_X^2 + D_X \nabla^2 C_X, \tag{4a}$$

$$\dot{C}_Y = -k_1 C_A C_H^2 C_Y - k_2 C_H C_X C_Y + \nu k_5 C_B C_Z$$
$$\quad + D_Y \nabla^2 C_Y, \tag{4b}$$

$$\dot{C}_Z = 2k_3 C_A C_H C_X - k_5 C_B C_Z + D_Z \nabla^2 C_Z, \tag{4c}$$

which describes the evolution of oscillating chemical concentrations C_X, C_Y, and C_Z for given supplied concentrations C_A, C_B, and C_H, and stoichiometric coefficient ν, which depends on the experimental setup. We vary the concentration of C_B and define a control parameter $\mu \equiv (C_B - C_B^c)/C_B^c$, where C_B^c is the critical value where the Hopf bifurcation occurs. Section S3A of the Supplemental Material details the Oregonator model, along with a data-driven approach to detect and characterize the bifurcation point when the model is unknown [30]. Here, we aim to develop a data-driven extension of Eq. (1) that incorporates nonlinear parameter dependence describing the dynamics far from the bifurcation.

We expect the dynamics near a Hopf point to be constrained to the two-dimensional center manifold, which describes the evolution of the complex amplitude dynamics governed by Eq. (1). When the governing equations are known, the weakly nonlinear theory develops a perturbative expansion near the Hopf point to express the complex amplitude A in terms of the state space given by $\mathbf{x} \equiv (C_X - C_X^c, C_Y - C_Y^c, C_Z - C_Z^c)$. This theory follows from a near identity transformation of the governing equations up to cubic order, as detailed in the Supplemental Material Sec. S3B [30].

To demonstrate our approach, we develop a data-driven construction of the amplitude A and its dynamical equations when the governing equations for the state space are unknown that is effective even far from the Hopf point. This approach follows from a SINDyCP fit on time series data for any two independent measurements of the state space (motivated by previous results [31], we choose to use C_X and C_Z here). Because we will employ the normal-form transformation below, we employ a cubic feature library with polynomial terms up to third-order and second-order spatial derivatives and a parameter library with polynomial terms up to second
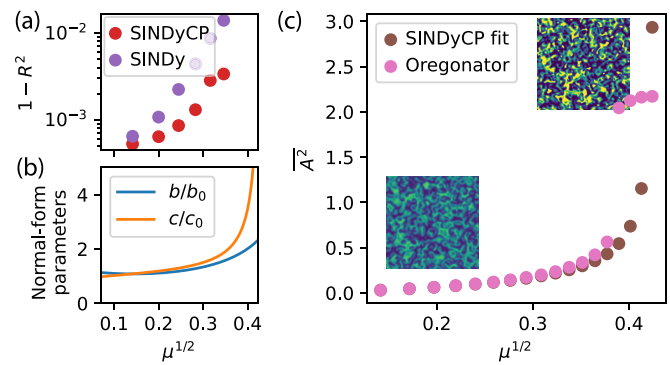

FIG. 2. Corrections to the weakly nonlinear theory of the Oregonator model. (a) R^2 score for the parameterized SINDyCP model and for an unparameterized SINDy fit on test trajectories collected at the parameter values used to train the model. (b) Corrected normal-form parameter values relative to the weakly nonlinear values b_0 and c_0 as a function of the bifurcation parameter $\mu^{1/2}$. (c) Average limit cycle amplitude $\overline{A^2}$ vs $\mu^{1/2}$ for the Oregonator model and SINDyCP fit. The fit correctly exhibits a canard explosion far from the onset of the instability (insets show C_X above and below the explosion, and an animation of the evolution is available in the Supplemental Material [30]).

order for the control parameter $\mu^{1/2}$. Furthermore, we employ implicit SINDy [32] by including first-order temporal derivatives in the feature library. Inverting the resulting implicit equations results in governing equations that are cubic in the state variables with coefficients that are rational functions of the control parameters, enabling the discovery of nonlinear corrections to parameter dependencies in the weakly nonlinear theory. Finally, by eliminating nonresonant coefficients using the normal-form transformation for cubic equations, we discover amplitude dynamics of the form in Eq. (1), but with normal-form coefficients $c(\mu)$ and $b(\mu)$ with rational dependence on the control parameter. Additional details on the data-driven amplitude construction are available in the Supplemental Material Sec. S3C [30].

Figure 2(a) shows the R^2 score of the model on test trajectories corresponding to the parameter values that the model was trained on (a value of $R^2 = 1$ means that the fit perfectly predicts the temporal derivatives of the data). For reference, we also perform an unparameterized SINDy fit [purple dots in Fig. 2(a)] on the training trajectory with the smallest μ value, which produces a model very close to the weakly nonlinear theory. The SINDyCP fit performs significantly better than the unparameterized fit, with $1 - R^2$ nearly an order of magnitude smaller for the larger μ values.

The normal-form parameters $b(\mu)$ and $c(\mu)$ agree with the analytic values derived [33] from the original model as $\mu \to 0$, but here we are able to discover them directly from data without any knowledge of the governing equations. Furthermore, as shown in Fig. 2(b), the variation of the parameters becomes extreme for $\mu^{1/2} > 0.35$, which we were able to discover via the implicit version of SINDy. In fact, as shown in Fig. 2(c), the Oregonator model exhibits a canard explosion (in which the limit cycle amplitude expands abruptly due to highly nonlinear effects) [31] around $\mu^{1/2} \approx 0.39$, where the weakly nonlinear theory breaks down. The SINDyCP model reflects

this breakdown and enables the development new models that account for it.

Weak formulation. The weak formulation utilizes integration against compactly supported "test functions" to defined the SINDy problem. The weak method shows excellent performance for noisy data, owing to its ability to minimize the need for computing numerical derivatives. Rather than forming samples (rows in Fig. 1) from spatiotemporal points for each trajectory, the weak method constructs the system rows by projecting the data onto weak samples such as

$$w_{ik}^\nu \equiv \int_{\Omega_k} \phi_k(\mathbf{x};t) X_i^{(\nu)}(\mathbf{x};t) d^D x dt, \tag{5}$$

where Ω_k is a compactly supported sample domain, ϕ_k is the test function, and $X_i^{(\nu)}$ denotes the νth partial derivative the ith dependent variable. By moving derivatives off of the data and onto the test functions via integration by parts,

$$w_{ik}^\nu = (-1)^{|\nu|} \int_{\Omega_k} \phi_k^{(\nu)}(\mathbf{x};t) X_i(\mathbf{x};t) d^D x dt, \tag{6}$$

the weak method significantly reduces the impact of measurement noise on the SINDy library and improves the fit results [34].

To maximize the performance of the weak method, we have optimized and fully vectorized numerical integration for the weak formulation in PySINDy, which can be easily combined with the SINDyCP library class. Products of weak features do not generally form reasonable samples for a SINDy model, since multiplication and integration do not commute, so at first sight, it is not clear how to combine weak-form feature and parameter libraries with SINDyCP. However, when computing the weak samples corresponding to constant functions, such as those that form the parameter library, the integrals simply represent the spatiotemporal volume of the domain Ω_k. Our implementation thus rescales the weak features for the temporal derivatives by the same volumetric factors. Details of our implementation are presented in Sec. S4 of the Supplemental Material [30].

Performance. Using 500 randomly distributed sample domains (measuring 1/10th the spatiotemporal domain size in each dimension), the weak SINDyCP easily identifies the complex Ginzburg-Landau equation using the same data used for the traditional differential form shown in Fig. 1. Furthermore, it can do so in just a few seconds of run time on a modern processor in this case (over five times faster than the differential form).

To assess the impact of noise, we inject random Gaussian noise of varying intensity [35] into the four trajectories used as the training data for the complex Ginzburg-Landau equation. We then generate two new sample trajectories to use as testing data, with $b = 2.0, 1.5$ and $c = 1.5, 1.0$, respectively. Using the training data, we perform the SINDyCP fits using both the differential and weak formulation and evaluate the R^2 score on our test trajectories. Figure 3(a) shows the results for the R^2 score on the test trajectories. While both methods provide good fits for low noise intensity, only the weak method exhibits a robust fit for parameterized equations for large noise intensities.

The SINDyCP fit also requires a sufficient amount of data to identify governing equations. Figure 3(b) shows the
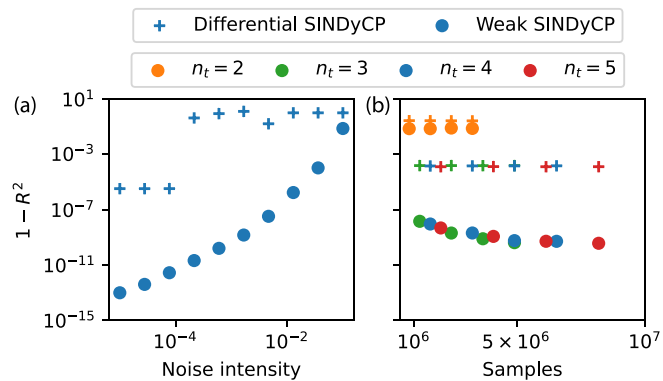

FIG. 3. Performance of SINDyCP for the fit of the complex Ginzburg-Landau equation with noisy data. (a) Model score vs noise intensity using the differential and weak forms of SINDyCP with $n_t = 4$ trajectories. (b) Model score vs number of samples for varying number of randomly generated trajectories, varying trajectory length, and noise intensity 10^{-3}.

performance of SINDyCP on the testing data for fits performed with a varying number of trajectories $n_t = 2, 3, 4, 5$ and of varying length corresponding to a number of time samples $N_t = 25, 50, 75, 100$, with an injected noise intensity of 10^{-3}. Unlike the trajectories in Fig. 1, the parameters for trajectories were randomly generated, with (b, c) distributed as Gaussian random variables with means $(1.5, 1.0)$ and standard deviations $(0.5, 0.25)$. For too little data, the fit fails to identify the correct model, and the value of $1 - R^2$ is $O(1)$. The models improve moderately with an increasing number of samples per trajectory (the product of N_t with the number of spatial grid points). More importantly, a sufficiently large number of trajectories n_t is required to achieve a good fit (at least 3 in this case). The amount of data required will further increase when including a larger number of possible library terms and when identifying a larger number of parameters. These requirements should be carefully assessed in order to achieve successful SINDyCP fits for more general pattern-forming systems.

Parameter extrapolation. As a final demonstration (Fig. 4), we consider the one-dimensional cubic-quintic Swift-Hohenberg equation

$$\dot{u} = du - u_{xxxx} - 2u_{xx} - u + eu^3 - fu^5, \tag{7}$$

with parameters d, e, and f describing the linear, cubic, and regularizing quintic terms, respectively. This model pattern-formation equation has been used to study defect dynamics incorporating quintic corrections beyond the weakly nonlinear approximation and has revealed universal snaking bifurcations leading to the formation of localized states for $e > 0$ and $d < 0$ [36].

The parameters d, e, and f are the minimal and natural set to describe the possible dynamics in the Swift-Hohenberg equation derived from normal-form theory. However, in typical pattern-formation applications, one does not have direct control over such parameters. Instead, experimentally accessible parameters will have a complicated and nonlinear relationship with the normal-form parameters, which requires detailed knowledge and tedious calculations to derive, e.g.,
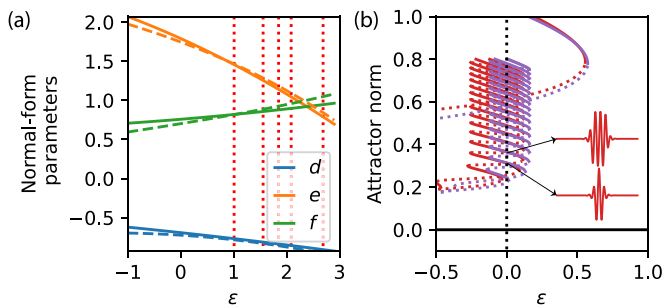
FIG. 4. Extrapolation of localized states in the cubic-quintic Swift-Hohenberg equation. (a) The randomly generated relationships between the normal-form parameters (d, e, f) and the experimental parameter ε. Red dotted lines show the values used to train the SINDyCP fit and dashed colored lines show the coefficients derived from the fit. (b) Snaking bifurcations of localized states for Eq. (7) (purple lines) and the SINDyCP fit (red lines). Insets show the localized states found from random initial conditions with $\varepsilon = 0$ in the discovered model.

an expansion and center manifold transformation around a bifurcation point. The SINDyCP approach enables an automated discovery of such relationships, which can be used to extrapolate system behavior beyond a set of measurements.

To illustrate this idea, we generate random quadratic relationships between an experimental parameter ε and the normal-form parameters (d, e, f), and we create five training trajectories using random values of the parameter $1 < \varepsilon < 3$ [Fig. 4(a)]. For all of the training trajectories, ε is sufficiently large that no localized or periodic states exist, and all trajectories decay to the trivial $u = 0$ solution. We perform the weak SINDyCP fit using these trajectories subject to injected white noise with intensity $\sigma = 0.01$ [35] and with a quadratic parameter library. To test the ability of SINDyCP to extrapolate beyond the parameter regime given in the input data, we simulate the identified model for the experimental parameter value $\varepsilon = 0$. Remarkably, even with limited and noisy training data, the method identifies an accurate relationship between ε and the normal-form parameters. Simulations of the identified model with random initial conditions converge to localized states for $\varepsilon = 0$. Numerical continuation of these localized states [Fig. 4(b)] with the AUTO package [37] reveals that the SINDyCP model exhibits snaking bifurcations that closely approximate those in the Swift-Hohenberg equation (see Sec. S5 of the Supplemental Material [30] for fits and continuations

with differing noise intensity and training data). Thus, despite the significant extrapolation of the parameter value beyond the input data, the model captures the complex bifurcation structure in the dynamics.

Discussion. The SINDyCP approach represents a simple but powerful generalization of SINDy with control. By disambiguating the feature and parameter components of the SINDy libraries, the method enables the discovery of systems of partial differential equations parameterized by driving parameters. Such equations arise naturally in the context of pattern formation, where the normal forms of bifurcations lead to parameterized equations near the onset of instabilities. The approach can be easily applied with the data available in typical pattern-formation experiments and promises to enable extrapolation beyond the regime that can be theoretically described with weakly nonlinear theory. For example, it may find application in the discovery of mechanisms leading to the formation of novel localized states beyond the snaking bifurcations of the Swift-Hohenberg equation [38,39]. While new phenomena may be easily conjectured to occur at unseen parameter values, we emphasize that such predictions must be validated experimentally to ensure correct extrapolation.

In practice, two significant challenges must be overcome to discover good parameterized models with SINDyCP. First, the method requires sufficiently informative trajectory data. Samples should be collected on appropriate temporal and spatial scales, sufficiently many parameter values should be measured, and trajectories with persistent dynamics provide more information than transient trajectories. While the weak formulation significantly mitigates the problem, measurement noise impairs the fit and can corrupt results. Second, the method requires an appropriate state space with a good coordinate representation to discover sparse dynamics. Near a bifurcation, the normal-form theory helps provide information about the state space dimension and sparsity-promoting coordinate transformations. In the future, a more sophisticated data-driven phase-amplitude reconstruction [40] or autoencoder-assisted discovery of physical coordinates [41–45] will further enable researchers to discover parsimonious equations governing complex systems directly from data gathered through experiments conducted under various driving parameters.

[1] S. L. Brunton and J. N. Kutz, *Data-Driven Science and Engineering: Machine Learning, Dynamical Systems, and Control* (Cambridge University Press, Cambridge, UK, 2022).

[2] S. M. Udrescu and M. Tegmark, AI Feynman: A physics-inspired method for symbolic regression, Sci. Adv. **6**, eaay2631 (2020).

[3] G. E. Karniadakis, I. G. Kevrekidis, L. Lu, P. Perdikaris, S. Wang, and L. Yang, Physics-informed machine learning, Nat. Rev. Phys. **3**, 422 (2021).

[4] S. L. Brunton, J. L. Proctor, and J. N. Kutz, Discovering governing equations from data by sparse identification of nonlinear dynamical systems, Proc. Natl. Acad. Sci. USA **113**, 3932 (2016).

[5] M. C. Cross and P. C. Hohenberg, Pattern formation outside of equilibrium, Rev. Mod. Phys. **65**, 851 (1993).

[6] Y. A. Kuznetsov, *Elements of Applied Bifurcation Theory* (Springer, New York, 1998).

[7] I. S. Aranson and L. Kramer, The world of the complex Ginzburg-Landau equation, Rev. Mod. Phys. **74**, 99 (2002).

[8] Z. G. Nicolaou, H. Riecke, and A. E. Motter, Chimera states in continuous media: Existence and distinctness, Phys. Rev. Lett. **119**, 244101 (2017).

[9] V. Heinonen, A. J. Abraham, J. Słomka, K. J. Burns, P. J. Sáenz, and J. Dunkel, Emergent universal statistics in nonequilibrium systems with dynamical scale selection, arXiv:2205.01627.

[10] S. L. Brunton, J. L. Proctor, and J. N. Kutz, Sparse identification of nonlinear dynamics with control (SINDYc), IFAC-PapersOnLine **49**, 710 (2016).

[11] E. Kaiser, J. N. Kutz, and S. L. Brunton, Sparse identification of nonlinear dynamics for model predictive control in the low-data limit, Proc. R. Soc. A **474**, 20180335 (2018).

[12] U. Fasel, E. Kaiser, J. N. Kutz, B. W. Brunton, and S. L. Brunton, SINDy with control: A tutorial, in *Proceedings of the 2021 60th IEEE Conference on Decision and Control (CDC)* (IEEE, Austin, 2021), pp. 16–21.

[13] S. H. Rudy, S. L. Brunton, J. L. Proctor, and J. N. Kutz, Data-driven discovery of partial differential equations, Sci. Adv. **3**, e1602614 (2017).

[14] H. Schaeffer, Learning partial differential equations via data discovery and sparse optimization, Proc. R. Soc. A **473**, 20160446 (2017).

[15] S. H. Rudy, A. Alla, S. Brunton, and J. N. Kutz, Data-driven identification of parametric partial differential equations, SIAM J. Appl. Dyn. Syst. **18**, 643 (2019).

[16] H. Schaeffer, G. Tran, and R. Ward, Learning dynamical systems and bifurcation via group sparsity, arXiv:1709.01558.

[17] B. Kramer, P. Grover, P. Boufounos, S. Nabi, and M. Benosman, Sparse sensing and DMD-based identification of flow regimes and bifurcations in complex flows, SIAM J. Appl. Dyn. Syst. **16**, 1164 (2017).

[18] M. Schmelzer, R. P. Dwight, and P. Cinnella, Discovery of algebraic Reynolds-stress models using sparse symbolic regression, Flow, Turbul. Combust. **104**, 579 (2020).

[19] L. Zanna and T. Bolton, Data-driven equation discovery of ocean mesoscale closures, Geophys. Res. Lett. **47**, e2020GL088376 (2020).

[20] S. Beetham, R. O. Fox, and J. Capecelatro, Sparse identification of multiphase turbulence closures for coupled fluid-particle flows, J. Fluid Mech. **914**, A11 (2021).

[21] Z. Wang, B. Wu, K. Garikipati, and X. Huan, A perspective on regression and Bayesian approaches for system identification of pattern formation dynamics, Theor. Appl. Mech. Lett. **10**, 188 (2020).

[22] N. Romeo, A. Hastewell, A. Mietke, and J. Dunkel, Learning developmental mode dynamics from single-cell trajectories, eLife **10**, e68679 (2021).

[23] R. Supekar, B. Song, A. Hastewell, G. P. T. Choi, A. Mietke, and J. Dunkel, Learning hydrodynamic equations for active matter from particle simulations and experiments, Proc. Natl. Acad. Sci. USA **120**, e2206994120 (2023).

[24] A. A. Kaptanoglu, B. M. de Silva, U. Fasel, K. Kaheman, A. J. Goldschmidt, J. L. Callaham, C. B. Delahunt, Z. G. Nicolaou, K. Champion, J. C. Loiseau, and J. N. Kutz, PySINDy: A comprehensive Python package for robust sparse system identification, J. Open Source Softw. **7**, 3994 (2022).

[25] The PySINDy repository is available at https://github.com/dynamicslab/pysindy.

[26] P. A. K. Reinbold, D. R. Gurevich, and R. O. Grigoriev, Using noisy or incomplete data to discover models of spatiotemporal dynamics, Phys. Rev. E **101**, 010203(R) (2020).

[27] P. A. K. Reinbold, L. M. Kageorge, M. F. Schatz, and R. O. Grigoriev, Robust learning from noisy, incomplete, high-dimensional experimental data via physically constrained symbolic regression, Nat. Commun. **12**, 3219 (2021).

[28] D. A. Messenger and D. M. Bortz, Weak SINDy for partial differential equations, J. Comput. Phys. **443**, 110525 (2021).

[29] D. A. Messenger and D. M. Bortz, Learning mean-field equations from particle data using WSINDy, Physica D **439**, 133406 (2022).

[30] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevResearch.5.L042017 for an animation of the trajectories in Figs. 1 and 2 and details about numerical integration, additional demonstrations, the Oregonator model, the weak-form implementation, and the Swift-Hohenberg equation.

[31] M. Mazzotti, M. Morbidelli, and G. Serravalle, Bifurcation analysis of the Oregonator model in the 3-D space bromate/malonic acid/stoichiometric coefficient, J. Phys. Chem. **99**, 4501 (1995).

[32] N. M. Mangan, S. L. Brunton, J. L. Proctor, and J. N. Kutz, Inferring biological networks by sparse identification of nonlinear dynamics, IEEE Trans. Mol. Biol. Multi-Scale Commun. **2**, 52 (2016).

[33] M. Ipsen, F. Hynne, and P. G. Sørensen, Amplitude equations for reaction-diffusion systems with a Hopf bifurcation and slow real modes, Physica D **136**, 66 (2000).

[34] It is not possible to remove all numerical derivatives in the weak formulation, but the maximum order of derivatives can generally be reduced to at most half the original order for the library.

[35] Noise intensity here refers to the pointwise standard deviation on the spatiotemporal grid employed in the simulations. True white noise has a Dirac delta variance, and intensity should thus scale with grid spacing and time step to 1/2 power.

[36] J. Burke and E. Knobloch, Homoclinic snaking: Structure and stability, Chaos **17**, 037102 (2007).

[37] AUTO is currently maintained in the GitHub repository, https://github.com/auto-07p/auto-07p.

[38] B. G. Chen, N. Upadhyaya, and V. Vitelli, Nonlinear conduction via solitons in a topological mechanical insulator, Proc. Natl. Acad. Sci. USA **111**, 13004 (2014).

[39] Z. G. Nicolaou, D. J. Case, E. B. Wee, M. M. Driscoll, and A. E. Motter, Heterogeneity-stabilized homogeneous states in driven media, Nat. Commun. **12**, 4486 (2021).

[40] B. Kralemann, L. Cimponeriu, M. Rosenblum, A. Pikovsky, and R. Mrowka, Phase dynamics of coupled oscillators reconstructed from data, Phys. Rev. E **77**, 066205 (2008).

[41] K. Champion, B. Lusch, J. N. Kutz, and S. L. Brunton, Data-driven discovery of coordinates and governing equations, Proc. Natl. Acad. Sci. USA **116**, 22445 (2019).

[42] T. E. Smidt, M. Geiger, and B. K. Miller, Finding symmetry breaking order parameters with Euclidean neural networks, Phys. Rev. Res. **3**, L012002 (2021).

[43] B. Chen, B. Huang, S. Raghupathi, I. Chandratreya, Q. Du, and H. Lipson, Automated discovery of fundamental variables hidden in experimental data, Nat. Comput. Sci. **2**, 433 (2022).

[44] J. Bakarji, K. Champion, J. N. Kutz, and S. L. Brunton, Discovering governing equations from partial measurements with deep delay autoencoders, Proc. R. Soc. A **479**, 20230422 (2023).

[45] M. Cenedese, J. Axås, B. Bäuerlein, K. Avila, and G. Haller, Data-driven modeling and prediction of non-linearizable dynamics via spectral submanifolds, Nat. Commun. **13**, 872 (2022).