APPLIED MATHEMATICS

Data-driven discovery of partial differential equations

Samuel H. Rudy,[1]* Steven L. Brunton,[2] Joshua L. Proctor,[3] J. Nathan Kutz[1]

We propose a sparse regression method capable of discovering the governing partial differential equation(s) of a given system by time series measurements in the spatial domain. The regression framework relies on sparsity-promoting techniques to select the nonlinear and partial derivative terms of the governing equations that most accurately represent the data, bypassing a combinatorially large search through all possible candidate models. The method balances model complexity and regression accuracy by selecting a parsimonious model via Pareto analysis. Time series measurements can be made in an Eulerian framework, where the sensors are fixed spatially, or in a Lagrangian framework, where the sensors move with the dynamics. The method is computationally efficient, robust, and demonstrated to work on a variety of canonical problems spanning a number of scientific domains including Navier-Stokes, the quantum harmonic oscillator, and the diffusion equation. Moreover, the method is capable of disambiguating between potentially nonunique dynamical terms by using multiple time series taken with different initial data. Thus, for a traveling wave, the method can distinguish between a linear wave equation and the Korteweg–de Vries equation, for instance. The method provides a promising new technique for discovering governing equations and physical laws in parameterized spatiotemporal systems, where first-principles derivations are intractable.

INTRODUCTION

Data-driven discovery methods, which have been enabled in the past decade by the plummeting cost of sensors, data storage, and computational resources, have a transformative impact on the sciences, facilitating a variety of innovations for characterizing high-dimensional data generated from experiments. Less well understood is how to uncover underlying physical laws and/or governing equations from time series data that exhibit spatiotemporal activity. Traditional theoretical methods for deriving the underlying partial differential equations (PDEs) are rooted in conservation laws, physical principles, and/or phenomenological behaviors. These first-principles derivations lead to many of the canonical models ubiquitous in physics, engineering, and the biological sciences. However, there remain many complex systems that have eluded quantitative analytic descriptions or even characterization of a suitable choice of variables (for example, neuroscience, power grids, epidemiology, finance, and ecology). We propose an alternative method to derive governing equations based solely on time series data collected at a fixed number of spatial locations. Using innovations in sparse regression, we discover the terms of the governing PDE that most accurately represent the data from a large library of potential candidate functions. Measurements can be made in an Eulerian framework, where the sensors are fixed spatially, or in a Lagrangian framework, where the sensors move with the dynamics. We demonstrate the success of the method by rediscovering a broad range of physical laws solely from time series data.

Methods for data-driven discovery of dynamical systems (*1*) include equation-free modeling (*2*), artificial neural networks (*3*), nonlinear regression (*4*), empirical dynamic modeling (*5*, *6*), normal form identification (*7*), nonlinear Laplacian spectral analysis (*8*), modeling emergent behavior (*9*), and automated inference of dynamics (*10*–*12*). In this series of developments, seminal contributions leveraging symbolic regression and an evolutionary algorithm (*13*, *14*) were capable of directly determining nonlinear dynamical system from data. More recently, sparsity (*15*) has been used to robustly determine, in a highly efficient computational manner, the governing dynamical system (*16*, *17*). Both the evolutionary (*14*) and sparse (*16*) symbolic regression methods avoid overfitting by selecting parsimonious models that balance model accuracy with complexity via Pareto analysis. The method we present is able to select, from a large library, the correct linear, nonlinear, and spatial derivative terms, resulting in the identification of PDEs from data. Only those terms that are most informative about the dynamics are selected as part of the discovered PDE. The innovation presented here is critically important because it efficiently handles spatiotemporal data, which is a fundamental characteristic of many canonical models. Previous sparsity-promoting methods are able to identify ordinary differential equations (ODEs) from data but are not able to handle spatiotemporal data or high-dimensional measurements (*16*). Our novel methodology has several advantageous practical characteristics: Measurements can be collected in either a fixed or moving frame (Eulerian or Lagrangian), allowing for a broad application to a variety of experimental data; the algorithm can also efficiently handle high-dimensional data through innovative sampling strategies. The algorithm, PDE functional identification of nonlinear dynamics (PDE-FIND), is applied to a wide range of canonical models.

RESULTS

We consider a parameterized and nonlinear PDE of the general form

$$u_t = N(u, u_x, u_{xx}, \cdots, x, \mu) \tag{1}$$

where the subscripts denote partial differentiation in either time or space, and $N(\cdot)$ is an unknown right-hand side that is generally a nonlinear function of $u(x, t)$, its derivatives, and parameters in μ. Our objective is to construct $N(\cdot)$ given time series measurements of the system at a fixed number of spatial locations in x. A key assumption is that the function $N(\cdot)$ consists of only a few terms, making the functional form sparse relative to the large space of possible contributing terms. As an example, Burgers' equation ($N = -uu_x + \mu u_{xx}$) and the harmonic oscillator ($N = -i\mu x^2 - i\hbar u_{xx}/2$) each have two terms.

[1]Department of Applied Mathematics, University of Washington, Seattle, WA 98195, USA. [2]Department of Mechanical Engineering, University of Washington, Seattle, WA 98195, USA. [3]Institute for Disease Modeling, 3150 139th Avenue Southeast, Bellevue, WA 98005, USA.
*Corresponding author. Email: shrudy@uw.edu

Thus, given the large collection of candidate terms for constructing PDEs, we use sparse regression methodologies to determine which right-hand-side terms are contributing to the dynamics without an intractable (np-hard) combinatorial brute-force search across all possible term combinations.

Upon discretization, the right hand side of Eq. 1 can be expressed as a function of $\mathbf{U}$, which is the discrete version of $u(x, t)$ and its derivatives, through the matrix $\boldsymbol{\Theta}(\mathbf{U}, \mathbf{Q})$, where the column vector $\mathbf{Q}$ contains any additional input terms to the right hand side. Each column of the library $\boldsymbol{\Theta}(\mathbf{U}, \mathbf{Q})$ corresponds to a specific candidate term for the governing equation, as shown in Fig. 1 (1b). The PDE evolution can be expressed in this library as follows

$$\mathbf{U}_t = \boldsymbol{\Theta}(\mathbf{U}, \mathbf{Q})\xi \tag{2}$$

Each nonzero entry in ξ corresponds to a term in the PDE, and for canonical PDEs, the vector ξ is sparse, meaning that only a few terms are active. We explicitly show in Materials and Methods how to construct $\boldsymbol{\Theta}(\mathbf{U}, \mathbf{Q})$ and solve for the vector ξ, thus identifying the terms in the PDE.

Discovering the Navier-Stokes equations

Figure 1 demonstrates the algorithmic procedure for successfully identifying the correct PDE dynamics for a given set of measurements from a physical system. Specifically, fluid flow around a cylinder is simulated at a given Reynolds, and measurements of the vorticity and velocity can be densely or sparsely sampled to correctly reconstruct the well-known Navier-Stokes equations. Remarkably, the coefficients of the PDE and Reynolds number are identified within a fraction of a percent accuracy. This figure represents our innovative mathematical structure that combines sparse regression, a library of potential functional forms, and parsimonious model selection.

Figure 1 also demonstrates that, for large data sets, such as those generated from two- and three-dimensional problems, PDE-FIND can be effectively used on subsampled data. This distinction is fundamentally important because full-state measurements are often computationally and experimentally prohibitive to collect and may also make the regression needlessly expensive. We randomly select a set of spatial points and uniformly subsample in time, resulting in the use of only a fraction of the data set. Mathematically, this amounts to ignoring a fraction of the rows in the linear system $\mathbf{U}_t = \boldsymbol{\Theta}(\mathbf{U}, \mathbf{Q})\xi$, as illustrated in Fig. 1 (2a and 2b). Although we only use a small fraction of the spatial points in the linear system, nearby points are needed to evaluate the derivative terms in the library. The derivatives are computed using a small number of spatially localized points near each measurement position via polynomial interpolation. Therefore, whereas subsampling uses only a small fraction of the points in the regression, we are using local information around each measurement.

Previous sparse identification algorithms (*16*) faced a number of challenges: They were not able to handle subsampled spatial data, and the algorithm did not scale well to high-dimensional measurements. Standard model reduction techniques such as proper orthogonal decomposition (POD) were used to overcome the high-dimensional measurements, allowing for a lower-order ODE model to be constructed on energetic POD modes. This procedure resembles the standard Galerkin projection onto POD modes (*18*). In contrast, the PDE-FIND algorithm identifies a PDE directly from subsampled measurement data.

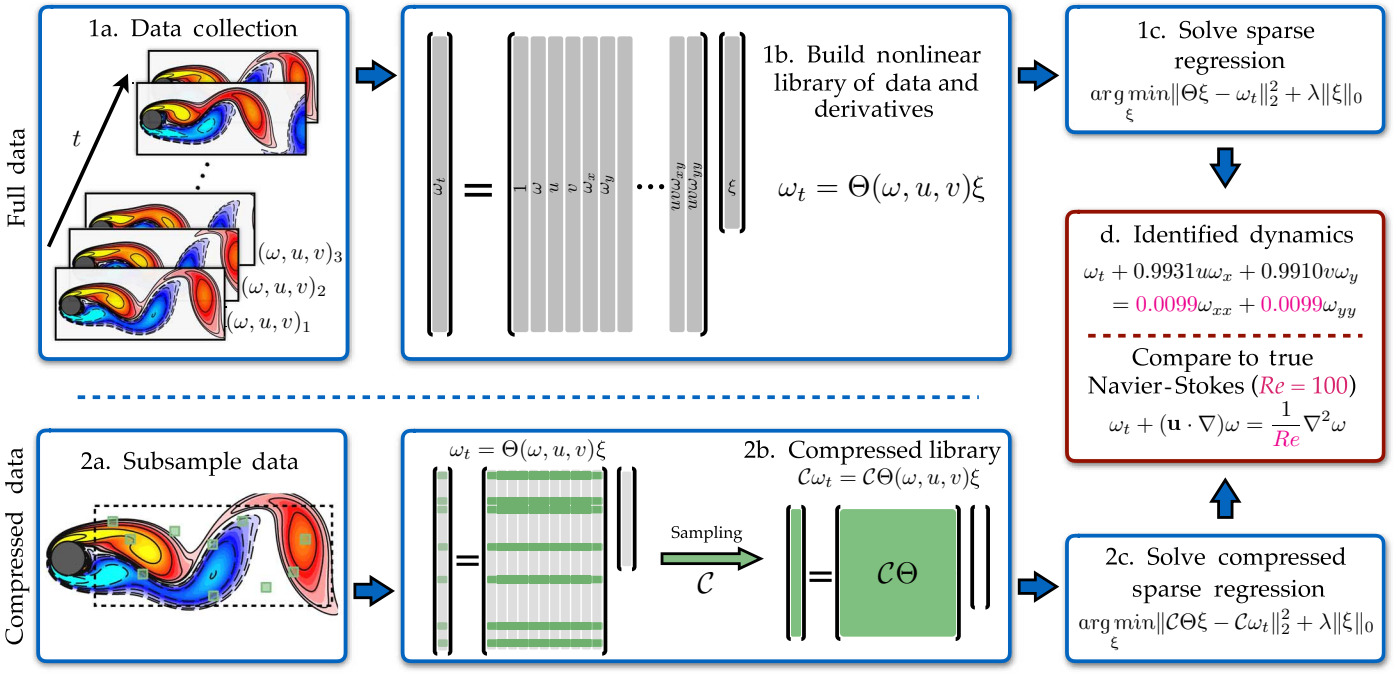


Fig. 1. Steps in the PDE functional identification of nonlinear dynamics (PDE-FIND) algorithm, applied to infer the Navier-Stokes equations from data. (1a) Data are collected as snapshots of a solution to a PDE. (1b) Numerical derivatives are taken, and data are compiled into a large matrix $\boldsymbol{\Theta}$, incorporating candidate terms for the PDE. (1c) Sparse regressions are used to identify active terms in the PDE. (2a) For large data sets, sparse sampling may be used to reduce the size of the problem. (2b) Subsampling the data set is equivalent to taking a subset of rows from the linear system in Eq. 2. (2c) An identical sparse regression problem is formed but with fewer rows. (d) Active terms in ξ are synthesized into a PDE.

Sensors moving with the dynamics

As a second demonstration of the method, we consider one of the fundamental results of the early 20th century concerning the relationship between random walks (Brownian motion) and diffusion. The theoretical connection between these two was first made by Einstein in 1905 (*19*) and was part of the Annus Mirabilis papers, which lay the foundations of modern physics. We use the method proposed here to sample the movement of a random walker, which is effectively a Lagrangian measurement coordinate, to verify that it can reproduce the well-known diffusion equation. By biasing the random walk, we can also produce the generalization of advection-diffusion in one dimension. Figure 2 shows the success of the method in identifying the correct diffusion model from a random walk trajectory. Given a sufficiently long time series with high enough resolution, the method produces the heat equation that describes the temporal evolution of the probability distribution function. It remains an open question how one might estimate the length of time series necessary to derive the underlying PDE. However, Fig. 2C suggests that a Pareto analysis, where the error drops sharply, may serve as a practical guide.

This example is important in that the measurements are now moving with the dynamics of the underlying governing equations. Specifically, measurements are made in the frame of the random walker. For many physical and engineering systems, fixed measurement locations may be impractical to achieve. This is relevant for applications such as ocean monitoring, where the change in spatial location of buoys due to ocean currents can be informative about the underlying dynamical properties of the system. These systems represent a Lagrangian framework for sensor placement and data-driven discovery of dynamics.

Disambiguating dynamical systems

A third canonical example is the Korteweg–de Vries (KdV) equation modeling the unidirectional propagation of small-amplitude, long or shallow-water waves. Discovered first by Boussinesq in 1877 and later developed by Korteweg and de Vries in 1895, it was one of the earliest models known to have soliton solutions. One potential viewpoint of the equation is as a dispersive regularization of Burgers' equation. The KdV evolution is given by

$$u_t + 6uu_x + u_{xxx} = 0 \tag{3}$$

with soliton solutions $u(x,t) = (c/2)\,\mathrm{sech}^2[(\sqrt{c}/2)(x - ct - x_0)]$, which propagate with a speed proportional to their amplitude c. If one observes a single propagating soliton, it would be indistinguishable from a solution to the one-way wave equation $u_t + cu_x = 0$. Hence, it presents a challenge to PDE-FIND, which would select the sparsest representation, in this case, the one-way wave equation. This ambiguity in selecting the correct PDE is rectified by constructing time series data for more than a single initial amplitude. Figure 3 demonstrates the evolution of two KdV solitons of differing amplitudes, which enables the unique determination of the governing PDE (Eq. 3).

Discovery for canonical PDE models

Table 1 applies the methodology proposed to a wide range of canonical models from mathematical physics and engineering sciences. The PDEs selected represent a wide range of physical systems, displaying both Hamiltonian (conservative) dynamics and dissipative nonlinear dynamics along with periodic to chaotic behavior. Aside from the quantum oscillator (third row), all the dynamics observed are strongly nonlinear. Each example is also identified after the addition of Gaussian noise to the data set, which significantly complicates accurate differentiation and library building. Remarkably, the method is able to discover each physical system even if significantly subsampled spatially. The space and time sampling used (data set in $\mathbb{C}^{n\times m}$), along with the accuracy in recovering the PDE parameters with and without noise, are detailed in Table 1. This highlights the broad applicability of the method and the success of the technique in discovering governing PDEs. Moreover, it demonstrates that a wide range of distinct physical observations can be captured by the regression framework, thus allowing the method to be broadly applied across the sciences, where spatiotemporal activity dominates.

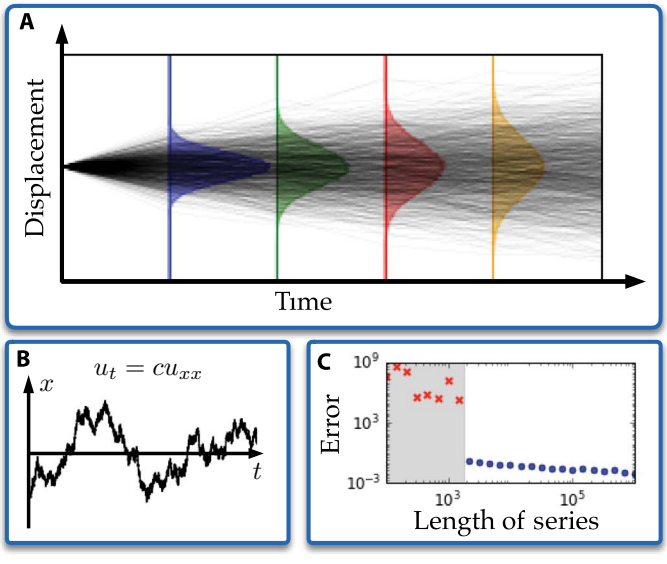


Fig. 2. Inferring the diffusion equation from a single Brownian motion. (**A**) Time series is broken into many short random walks that are used to construct histograms of the displacement. (**B**) Brownian motion trajectory following the diffusion equation. (**C**) Parameter error ($||\hat{\xi} - \xi * ||_1$) versus length of known time series. Blue symbols correspond to correct identification of the structure of the diffusion model $u_t = cu_{xx}$.

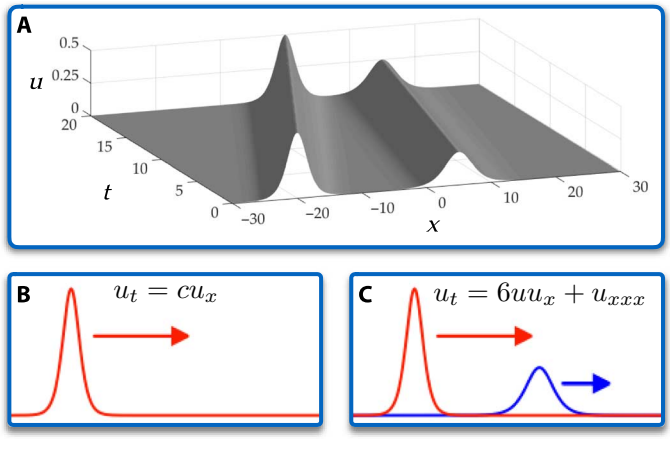
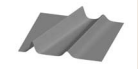
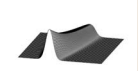
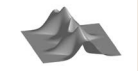
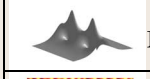
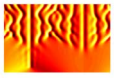
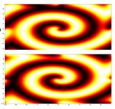


Fig. 3. Inferring nonlinearity via observing solutions at multiple amplitudes. (**A**) Example two-soliton solution to the KdV equation. (**B**) Applying our method to a single soliton solution determines that it solves the standard advection equation. (**C**) Looking at two completely separate solutions reveals nonlinearity.

Table 1. Summary of regression results for a wide range of canonical models of mathematical physics. In each example, the correct model structure is identified using PDE-FIND. The spatial and temporal sampling of the numerical simulation data used for the regression is given along with the error produced in the parameters of the model for both no noise and 1% noise. In the reaction-diffusion system, 0.5% noise is used. For Navier-Stokes and reaction-diffusion, the percent of data used in subsampling is also given. NLS, nonlinear Schrödinger; KS, Kuramoto-Sivashinsky.

PDE	Form	Error (no noise, noise)	Discretization
 KdV	$u_t + 6uu_x + u_{xxx} = 0$	$1 \pm 0.2\%, 7 \pm 5\%$	$x \in [-30, 30], n = 512,\ t \in [0, 20], m = 201$
 Burgers	$u_t + uu_x - \epsilon u_{xx} = 0$	$0.15 \pm 0.06\%, 0.8 \pm 0.6\%$	$x \in [-8, 8], n = 256,\ t \in [0, 10], m = 101$
 Schrödinger	$iu_t + \frac{1}{2}u_{xx} - \frac{x^2}{2}u = 0$	$0.25 \pm 0.01\%, 10 \pm 7\%$	$x \in [-7.5, 7.5], n = 512,\ t \in [0, 10], m = 401$
 NLS	$iu_t + \frac{1}{2}u_{xx} + \lvert u\rvert^2 u = 0$	$0.05 \pm 0.01\%, 3 \pm 1\%$	$x \in [-5, 5], n = 512,\ t \in [0, \pi], m = 501$
 KS	$u_t + uu_x + u_{xx} + u_{xxxx} = 0$	$1.3 \pm 1.3\%, 52 \pm 1.4\%$	$x \in [0, 100], n = 1024,\ t \in [0, 100], m = 251$
u  v Reaction Diffusion	$u_t = 0.1\nabla^2 u + \lambda(\mathrm{A})u - \omega(\mathrm{A})v$ $v_t = 0.1\nabla^2 v + \omega(\mathrm{A})u + \lambda(\mathrm{A})v$ $\mathrm{A}^2 = u^2 + v^2, \omega = -\beta\,\mathrm{A}^2, \lambda = 1 - \mathrm{A}^2$	$0.02 \pm 0.01\%,\ 3.8 \pm 2.4\%$	$x, y \in [-10, 10], n = 256,\ t \in [0, 10], m = 201$ subsample 1.14%
 Navier-Stokes	$\omega_t + (\mathbf{u} \cdot \nabla)\omega = \frac{1}{Re}\nabla^2\omega$	$1 \pm 0.2\%\ ,\ 7 \pm 6\%$	$x \in [0, 9], n_x = 449,\ y \in [0, 4], n_y = 199,$ $t \in [0, 30], m = 151,$ subsample 2.22%

DISCUSSION

PDE-FIND is a viable, data-driven tool for modern applications, where first-principles derivations may be intractable (for example, neuroscience, epidemiology, and dynamical networks) but where new data recordings and sensor technologies are revolutionizing our understanding of physical and/or biophysical processes on spatial domains. To our knowledge, this is the first data-driven regression technique that explicitly accounts for spatial derivatives in discovering physical laws, thus significantly broadening the applicability to a wide variety of complex systems. Traditional approaches, such as reducing the model to a local linear embedding based on collected data, fall short in achieving the fundamental goal of research scientists: to construct a nonlinear model from observations, which characterizes the observed dynamics and generalizes to unsampled areas of parameter space. For instance, we can discover the Navier-Stokes equations at $Re = 100$ and a particular boundary condition and use this knowledge to accurately simulate new boundary conditions where no data were collected. As in the study of Brunton *et al.* (*16*), it may also be possible to sample the dynamics at a range of parameter values and discover the fully parameterized system. The goal would be to collect data at a range of flow conditions corresponding to different Reynolds numbers, and infer both the parameterized Navier-Stokes equations and the Reynolds number, which is an exciting avenue of future research.

The Kuramoto-Sivashinsky equation is particularly challenging to identify with low error on the coefficients and illustrates a limitation of the method. Although the correct terms are identified, there is substantial coefficient error in the case with noise. The true and identified models are given by

$$u_t + uu_x + u_{xx} + u_{xxxx} = 0 \qquad \text{(True)}$$
$$u_t + 0.46uu_x + 0.48u_{xx} + 0.49u_{xxxx} = 0 \qquad \text{(Identified)}$$

The high precision but substantial underestimation of the coefficients indicates significant error in numerical differentiation.

The success of this methodology suggests that many concepts from statistical learning can be integrated with traditional scientific computing and dynamical systems theory to discover dynamical models from data. This integration of nonlinear dynamics and machine learning opens the door for principled versus heuristic methods for model construction, nonlinear control strategies, and sensor placement techniques.

MATERIALS AND METHODS

The innovations presented here were built upon three key ideas: (i) overcomplete libraries of candidate functions to represent the dynamics, (ii) sparse regression for selecting a small number of terms, and (iii) parsimonious selection of the governing equations via Pareto analysis. The architecture is enabled by the ability to accurately compute derivatives even when noise is present in the measurements. Each method is detailed below.

Building libraries of candidate terms

The sparse regression and discovery method (see Fig. 1) begins by first collecting all the spatial time series data into a single column vector $\mathbf{U} \in \mathbb{C}^{n \times m}$, representing data collected over m time points and n spatial locations. We also consider any additional input such as a known potential for the Schrödinger equation, or the magnitude of complex

data, in a column vector $\mathbf{Q} \in \mathbb{C}^{n \times m}$. Next, a library $\mathbf{\Theta}(\mathbf{U}, \mathbf{Q}) \in \mathbb{C}^{nm \times D}$ of D candidate linear and nonlinear terms and partial derivatives for the PDE is constructed. Derivatives are taken either using finite differences for clean data or when noise is added with polynomial interpolation. The candidate linear and nonlinear terms and partial derivatives are then combined into a matrix $\mathbf{\Theta}(\mathbf{U}, \mathbf{Q})$, which takes the form, for example

$$\mathbf{\Theta}(\mathbf{U}, \mathbf{Q}) = [1 \;\; \mathbf{U} \;\; \mathbf{U}^2 \;\; \ldots \;\; \mathbf{Q} \;\; \ldots \;\; \mathbf{U}_x \;\; \mathbf{U}\mathbf{U}_x \;\; \ldots] \tag{4}$$

Each column of $\mathbf{\Theta}$ contains all of the values of a particular candidate function across all of the mn space-time grid points on which data are collected. Therefore, if we have data on an $n \times m$ grid (for example, a 256 × 100 grid represents 256 spatial measurements at 100 time points) and have 50 candidate terms in the PDE, then $\mathbf{\Theta} \in \mathbb{C}^{256 \cdot 100 \times 50}$. The time derivative U_t was also computed and reshaped into a column vector. Figure 1 demonstrates the data collection and processing. As an example, a column of $\mathbf{\Theta}(\mathbf{U}, \mathbf{Q})$ may be $q^2 u_{xx}$.

If we assume that $\mathbf{\Theta}$ is an overcomplete library, meaning $\mathbf{\Theta}$ has a sufficiently rich column space that the dynamics will be in its range, then the PDE should be well represented by Eq. 2 with a sparse vector of coefficients ξ. This is equivalent to picking enough candidate functions that the full PDE may be written as a weighted sum of library terms. For an unbiased representation of the dynamics, we would simply solve the least-squares problem for ξ. However, even with the only error coming from numerical roundoff, the least-squares solution may be inaccurate. In particular, ξ will have predominantly nonzero values, suggesting a PDE with every functional form contained in the library. Regression problems similar to Eq. 2 result in least-squares problems that are poorly conditioned. Error in computing the derivatives (already an ill-conditioned problem with noise) will be magnified by numerical errors when inverting $\mathbf{\Theta}$. Thus, a least-squares regression radically changes the qualitative nature of the inferred dynamics.

Sparse regression

In general, we require the sparsest vector ξ that satisfies Eq. 2 with a small residual. Instead of an intractable combinatorial search through all possible sparse vector structures, a common technique was to relax the problem to a convex ℓ^1 regularized least squares (*15*); however, this tends to perform poorly with highly correlated data. Instead, we approximate the problem using candidate solutions to a ridge regression problem with hard thresholding, which we call sequential threshold ridge regression (STRidge in Algorithm 1). For a given tolerance and λ, this gives a sparse approximation to ξ. We iteratively refined the tolerance of Algorithm 1 to find the best predictor based on the selection criteria

$$\hat{\xi} = \underset{\xi}{\operatorname{argmin}} \, ||\mathbf{\Theta}(\mathbf{U}, \mathbf{Q})\xi - \mathbf{U}_t||_2^2 + \epsilon\kappa(\mathbf{\Theta}(\mathbf{U}\mathbf{Q}))||\xi||_0 \tag{5}$$

where $\kappa(\mathbf{\Theta})$ is the condition number of the matrix $\mathbf{\Theta}$, indicating stronger regularization for ill-posed problems. Penalizing $\|\xi\|_0$ discourages overfitting by selecting from the optimal position in a Pareto front.

Other methods for finding sparse solutions to least-squares problem exist. Greedy algorithms were shown to exhibit good performance on sparse optimization problems including PDE-FIND but, in some cases, were less reliable than STRidge (*20*). Although STRidge with normalization works well on almost all of the examples we tested, we do not make the claim that it is optimal. If additional information regarding the PDE is known (for instance, if we know that one of the terms is nonzero), then this may be incorporated into the penalty on the coefficients.

Numerical evaluation of derivatives

Proper evaluation of the numerical derivatives is the most challenging and critical task for the success of the method. This is particularly true when the discretized solution contains measurement noise. Given the well-known accuracy problems with finite-difference approximations, we experimented with a number of more robust numerical differentiation methods. Smoothing with a Gaussian kernel and Tikhonov differentiation were both investigated but were difficult to implement because of the difficulty in controlling the bias-variance trade-off. Spectral differentiation with thresholding for high-frequency terms was also considered but not used because of its restriction to periodic domains and the difficulty in implementing an appropriate thresholding function. The most reliable and robust method for computing derivatives from noisy data was polynomial interpolation (*21*). For each data point where we compute a derivative, a polynomial of degree P was fit to greater than P points, and derivatives of the polynomial were taken to approximate those of the numerical data. Points close to the boundaries, where it was difficult to fit a polynomial, were not used in the regression. This method is far from perfect; data close to the boundaries were difficult to differentiate, and we discovered that this could strongly influence the results of PDE-FIND. For a more principled but involved approach to polynomial differentiation, see the study of Bruno and Hoch (*22*).

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at http://advances.sciencemag.org/cgi/content/full/3/4/e1602614/DC1

Introduction
PDE-FIND
Examples
Limitations
fig. S1. Steps in the PDE functional identification of nonlinear dynamics (PDE-FIND) algorithm, applied to infer the Navier-Stokes equations from data.
fig. S2. The numerical solution to the KdV equation plotted in space-time.
fig. S3. The numerical solution to the Burgers' equation plotted in space-time.
fig. S4. The magnitude of the numerical solution to the Schrödinger's equation plotted in space-time.
fig. S5. The magnitude of the numerical solution to the nonlinear Schrödinger's equation plotted in space-time.
fig. S6. The numerical solution to the Kuramoto-Sivashinsky equation plotted in space-time.
fig. S7. The numerical solution to the reaction-diffusion equation plotted in space-time.
fig. S8. A single snapshot of the vorticity field is illustrated for the fluid flow past a cylinder.
fig. S9. A single stochastic realization of Brownian motion.
fig. S10. Five empirical distributions, illustrating the statistical spread of a particle's expected location over time, are presented.
fig. 11. Five empirical distributions, illustrating the statistical spread of a particle's expected location over time, are presented.
fig. S12. The numerical solution to the misidentified Kuramoto-Sivashinsky equation.
fig. S13. The numerical solution to the misidentified nonlinear Schrödinger equation.
fig. S14. Results of PDE-FIND applied to Burgers' equation for varying levels of noise.
table S1. Summary of regression results for a wide range of canonical models of mathematical physics.
table S2. Summary of PDE-FIND for identifying the KdV equation.
table S3. Summary of PDE-FIND for identifying Burgers' equation.
table S4. Summary of PDE-FIND for identifying the Schrödinger equation.
table S5. Summary of PDE-FIND for identifying the nonlinear Schrödinger equation.
table S6. Summary of PDE-FIND for identifying the Kuramoto-Sivashinsky equation.
table S7. Summary of PDE-FIND for identifying reaction-diffusion equation.

table S8. Summary of PDE-FIND for identifying the Navier-Stokes equation.
table S9. Accuracy of PDE-FIND on Burgers' equation with various grid sizes.
References (*23*–*50*)

REFERENCES AND NOTES

1. J. Crutchfield, B. McNamara, Equations of motion from a data series. *Complex Syst.* **1**, 417–452 (1987).
2. C. W. Gear, J. M. Hyman, P. G. Kevrekidis, I. G. Kevrekidis, O. Runborg, C. Theodoropoulos, Equation-free, coarse-grained multiscale computation: Enabling mocroscopic simulators to perform system-level analysis. *Commun. Math. Sci.* **1**, 715–762 (2003).
3. R. González-García, R. Rico-Martínez, I. G. Kevrekidis, Identification of distributed parameter systems: A neural net based approach. *Comput. Chem. Eng.* **22**, S965–S968 (1998).
4. H. U. Voss, P. Kolodner, M. Abel, J. Kurths, Amplitude equations from spatiotemporal binary-fluid convection data. *Phys. Rev. Lett.* **83**, 3422 (1999).
5. G. Sugihara, R. May, H. Ye, C.-h. Hsieh, E. Deyle, M. Fogarty, S. Munch, Detecting causality in complex ecosystems. *Science* **338**, 496–500 (2012).
6. H. Ye, R. J. Beamish, S. M. Glaser, S. C. H. Grant, C.-h. Hsieh, L. J. Richards, J. T. Schnute, G. Sugihara, Equation-free mechanistic ecosystem forecasting using empirical dynamic modeling. *Proc. Natl. Acad. Sci. U.S.A.* **112**, E1569–E1576 (2015).
7. A. J. Majda, C. Franzke, D. Crommelin, Normal forms for reduced stochastic climate models. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 3649–3653 (2009).
8. D. Giannakis A. J. Majda, Nonlinear Laplacian spectral analysis for time series with intermittency and low-frequency variability. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 2222–2227 (2012).
9. A. J. Roberts, *Model Emergent Dynamics in Complex Systems* (SIAM, 2014).
10. M. D. Schmidt, R. R. Vallabhajosyula, J. W. Jenkins, J. E. Hood, A. S. Soni, J. P. Wikswo, H. Lipson, Automated refinement and inference of analytical models for metabolic networks. *Phys. Biol.* **8**, 055011 (2011).
11. B. C. Daniels I. Nemenman, Automated adaptive inference of phenomenological dynamical models. *Nat. Commun.* **6**, 8133 (2015).
12. B. C. Daniels I. Nemenman, Efficient inference of parsimonious phenomenological models of cellular dynamics using S-systems and alternating regression. *PLOS ONE* **10**, e0119821 (2015).
13. J. Bongard H. Lipson, Automated reverse engineering of nonlinear dynamical systems. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 9943–9948 (2007).
14. M. Schmidt H. Lipson, Distilling free-form natural laws from experimental data. *Science* **324**, 81–85 (2009).
15. R. Tibshirani, Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. B* **58**, 267–288 (1996).
16. S. L. Brunton, J. L. Proctor, J. N. Kutz, Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 3932–3937 (2016).
17. N. M. Mangan, S. L. Brunton, J. L. Proctor, J. N. Kutz, Inferring biological networks by sparse identification of nonlinear dynamics. *IEEE Trans. Mol. Biol. Multi-Scale Commun.* **2**, 52–63 (2016).
18. P. Holmes, J. L. Lumley, G. Berkooz, C. W. Rowley, *Turbulence, Coherent Structures, Dynamical Systems and Symmetry* (Cambridge Univ. Press, ed. 2, 2012).
19. A. Einstein, Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen. *Ann. Phys.* **322**, 549–560 (1905).
20. T. Zhang, Adaptive forward-backward greedy algorithm for sparse learning with linear models, in *Advances in Neural Information Processing Systems 21*, D. Koller, D. Schuurmans, Y. Bengio, L. Bottou, Eds. (Curran Associates Inc., 2009), pp. 1921–1928.
21. I. Knowles, R. J. Renka, Methods for numerical differentiation of noisy data. *Electron. J. Differ. Eq.* 235–246 (2014).
22. O. Bruno, D. Hoch, Numerical differentiation of approximated functions with limited order-of-accuracy deterioration. *SIAM J. Numer. Anal.* **50**, 1581–1603 (2012).
23. Z. Bai, T. Wimalajeewa, Z. Berger, G. Wang, M. Glauser, P. K. Varshney, Low-dimensional approach for reconstruction of airfoil data via compressive sensing. *AIAA J.* **53**, 920–933 (2014).
24. G. Berkooz, P. Holmes, J. L. Lumley, The proper orthogonal decomposition in the analysis of turbulent flows. *Annu. Rev. Fluid Mech.* **25**, 539–575 (1993).
25. S. L. Brunton, B. W. Brunton, J. L. Proctor, E. Kaiser, J. N. Kutz, Chaos as an intermittently forced linear system. arXiv:1608.05306 (2016).
26. S. L. Brunton, J. H. Tu, I. Bright, J. N. Kutz, Compressive sensing and low-rank libraries for classification of bifurcation regimes in nonlinear dynamical systems. *SIAM J. Appl. Dyn. Syst.* **13**, 1716–1732 (2014).
27. M. Budišić, R. Mohr, I. Mezić, Applied Koopmanism. *Chaos* **22**, 047510 (2012).
28. J. D. Cole, On a quasi-linear parabolic equation occurring in aerodynamics. *Quart. Appl. Math.* **9**, 225–236 (1951).
29. T. Colonius, K. Taira, A fast immersed boundary method using a nullspace approach and multi-domain far-field boundary conditions. *Comput. Methods Appl. Mech. Eng.* **197**, 2131–2146 (2008).
30. M. C. Cross, P. C. Hohenberg, Pattern formation out of equilibrium. *Rev. Mod. Phys.* **65**, 851–1112 (1993).
31. C. Gardiner, *Stochastic Methods* (Springer, 2009).
32. M. Gavish, D. L. Donoho, The optimal hard threshold for singular values is $4/\sqrt{3}$. *IEEE Trans. Inf. Theory* **60**, 5040–5053 (2014).
33. E. Hopf, The partial differential equation $u_t + uu_x = \mu u_{xx}$. *Commun. Pure App. Math.* **3**, 201–230 (1950).
34. J. N. Kutz, *Data-Driven Modeling & Scientific Computation: Methods for Complex Systems & Big Data* (Oxford Univ. Press, 2013).
35. R. J. LeVeque, *Finite Difference Methods for Ordinary and Partial Differential Equations: Steady-State and Time-Dependent Problems, Vol. 98* (SIAM, 2007).
36. A. Mackey, H. Schaeffer, S. Osher, On the compressive spectral method. *Multiscale Model. Simul.* **12**, 1800–1827 (2014).
37. I. Mezić, Spectral properties of dynamical systems, model reduction and decompositions. *Nonlinear Dyn.* **41**, 309–325 (2005).
38. I. Mezić, Analysis of fluid flows via spectral properties of the Koopman operator. *Annu. Rev. Fluid Mech.* **45**, 357–378 (2013).
39. K. P. Murphy, *Machine Learning: A Probabilistic Perspective* (MIT Press, 2012).
40. B. R. Noack, K. Afanasiev, M. Morzyński, G. Tadmor, F. Thiele, A hierarchy of low-dimensional models for the transient and post-transient cylinder wake. *J. Fluid Mech.* **497**, 335–363 (2003).
41. V. Ozoliņš, R. Lai, R. Caflisch, S. Osher, Compressed modes for variational problems in mathematics and physics. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 18368–18373 (2013).
42. R. Pinnau, Model reduction via proper orthogonal decomposition, in *Model Order Reduction: Theory, Research Aspects and Applications* (Springer, 2008) pp. 95–109.
43. J. L. Proctor, S. L. Brunton, B. W. Brunton, J. N. Kutz, Exploiting sparsity and equation- free architectures in complex systems. *Eur. Phys. J. Spec. Top.* **223**, 2665–2684 (2014).
44. H. Schaeffer, R. Caflisch, C. D. Hauck, S. Osher, Sparse dynamics for partial differential equations. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 6634–6639 (2013).
45. K. Taira, T. Colonius, The immersed boundary method: A projection approach. *J. Comput. Phys.* **225**, 2118–2137 (2007).
46. G. Tran, R. Ward, Exact recovery of chaotic systems from highly corrupted data. arXiv:1607.01067 (2016).
47. L. N. Trefethen, *Spectral Methods in MATLAB, Vol. 10* (SIAM, 2000).
48. L. N. Trefethen, D. Bau III, *Numerical Linear Algebra, Vol. 50* (SIAM, 1997).
49. W.-X. Wang, R. Yang, Y.-C. Lai, V. Kovanis, C. Grebogi, Predicting catastrophes in non-linear dynamical systems by compressive sensing. *Phys. Rev. Lett.* **106**, 154101 (2011).
50. H. Zou, T. Hastie, Regularization and variable selection via the elastic net. *J. R. Stat. Soc. B* **67**, 301–320 (2005).

Acknowledgments

Funding: J.N.K. acknowledges support from the Air Force Office of Scientific Research (FA9550-15-1-0385). S.L.B. and J.N.K. acknowledge support from the Defense Advanced Research Projects Agency (DARPA contract HR0011-16-C-0016). J.L.P. would like to thank Bill and Melinda Gates for their active support of the Institute for Disease Modeling and their sponsorship through the Global Good Fund. **Author contributions:** S.H.R. developed the sparse selection algorithms. S.H.R. and J.N.K. developed simulation codes for the example problems considered. All authors participated in designing the study and discussing and writing the results. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Additional data related to this paper may be requested from the authors. All data and codes used in this manuscript are publicly available on GitHub at https://github.com/snagcliffs/PDE-FIND.

Submitted 23 October 2016
Accepted 11 February 2017
Published 26 April 2017
10.1126/sciadv.1602614

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