Special Issue on Biological Applications of Information Theory in Honor of Claude Shannon's Centennial—Part 1

Inferring Biological Networks by Sparse Identification of Nonlinear Dynamics

Niall M. Mangan, Steven L. Brunton, *Member, IEEE*, Joshua L. Proctor, and J. Nathan Kutz, *Member, IEEE*

***Abstract*—Inferring the structure and dynamics of network models is critical to understanding the functionality and control of complex systems, such as metabolic and regulatory biological networks. The increasing quality and quantity of experimental data enable statistical approaches based on information theory for model selection and goodness-of-fit metrics. We propose an alternative data-driven method to infer networked nonlinear dynamical systems by using sparsity-promoting optimization to select a subset of nonlinear interactions representing dynamics on a network. In contrast to standard model selection methods-based upon information content for a finite number of heuristic models (order 10 or less), our model selection procedure discovers a parsimonious model from a combinatorially large set of models, without an exhaustive search. Our particular innovation is appropriate for many biological networks, where the governing dynamical systems have rational function nonlinearities with cross terms, thus requiring an implicit formulation and the equations to be identified in the null-space of a library of mixed nonlinearities, including the state and derivative terms. This method, implicit-SINDy, succeeds in inferring three canonical biological models: 1) Michaelis–Menten enzyme kinetics; 2) the regulatory network for competence in bacteria; and 3) the metabolic network for yeast glycolysis.**

***Index Terms*—Dynamical systems, network inference, nonlinear dynamics, biological networks, machine learning, sparse selection, non-convex optimization.**

I. Introduction

NETWORK science is of growing and critical importance across the physical, biological and engineering sciences. In biology, both the quality and quantity of modern data has inspired new mathematical techniques for inferring the complex interactions and connections between nodes in metabolic and regulatory networks (see [1]–[5] and references therein). Discovering the structure, connectivity and dynamics of such networks is critical in understanding the functionality and control decisions enacted by the network in tasks such as cell differentiation, cell death, or directing metabolic flux [6]–[9]. Methods based on information theory provide rigorous statistical criteria for such model selection and network inference tasks. For example, partnering the Kullback-Leibler (KL) divergence [10], [11], a measure of information loss between empirically collected data and model generated data, and the Akaike information criteria (AIC) [12], [13], a relative estimate of information loss across models balancing model complexity and goodness-of-fit, allows for a principled model selection criteria [14]. Indeed, model selection is a fundamentally important computation for a variety of areas of biology such as epidemiology, neuroscience, and gene networks, where first principles models are unavailable.

Model selection methods and information theoretic approaches for nonlinear dynamical networks, however, face significant challenges. To successfully formulate dynamic models, a method must determine both the correct nodal dynamics and its node-to-node connectivity from a combinatorial large space of potential interactions. Traditional model selection is performed using information criteria on a finite, generally order 10 or less, set of heuristically defined models [15]–[18]. The model with the maximal information content is selected using AIC [12], Bayesian information criteria (BIC) [19], or other related metrics [20]–[23].

A continued goal of model selection is to consider a more comprehensive set of models, potentially combinatorially large, for evaluation by information criteria. Recent efforts significantly enhance the number of potential models by automated model generation [24], [25]. Despite the significant increase in candidate models, even a modest number of interacting variables can render model selection algorithms intractable. This is because each model posited must be numerically simulated, fit and evaluated for information content. To highlight this complexity, consider selecting a specific model from all possible polynomials of degree d in n variables. The number of possible monomial structures N_m is given by $N_m = \binom{n+d}{d}$, while the number of polynomial structures, N_p, that may be formed by assigning nonzero coefficients to the N_m monomials is given by: $N_p = \sum_{k=1}^{N_m} \binom{N_m}{k}$. The number of possible polynomial structures N_p may be thought of as summing over all possible polynomials with only k monomials. In our metabolic network example below, we consider

Manuscript received May 24, 2016; revised September 15, 2016 and November 16, 2016; accepted November 21, 2016. Date of current version January 6, 2017. The work of S. L. Brunton and J. N. Kutz was supported by the Defense Advanced Research Projects Agency under Contract HR0011-16-C-0016. The associate editor coordinating the review of this paper and approving it for publication was P. J. Thomas.

N. M. Mangan and J. N. Kutz are with the Department of Applied Mathematics, University of Washington, Seattle, WA 98195 USA (e-mail: kutz@uw.edu).

S. L. Brunton is with the Department of Mechanical Engineering, University of Washington, Seattle, WA 98195 USA, and also with the eScience Institute, University of Washington, Seattle, WA 98195 USA.

J. L. Proctor is with the Institute for Disease Modeling, Bellevue, WA 98005 USA.

Digital Object Identifier 10.1109/TMBMC.2016.2633265

polynomials up to degree $d = 4$ with $n = 5$ variables which results in $N_m = 126$ and $N_p \approx 10^{38}$. Simulating, evaluating, and comparing 10^{38} models in a traditional model selection framework using information criteria is currently intractable.

To overcome the significant challenge of model selection with a combinatorially large set of models, we propose a novel method to infer both the dynamics and connectivity of biological networks motivated by machine learning methods, overcomplete libraries [26]–[28], and non-convex optimization [29]. We demonstrate the accuracy and robustness of the method, called *implicit sparse identification of nonlinear dynamics* (implicit-SINDy), on three representative dynamic and networked biological models. Our combinatorial selection criteria is related to information criteria through a comparison of the Pareto analysis of implicit-SINDy to a likelihood function that minimizes the difference between the model and data. The Pareto front discovers the most parsimonious model, similar to AIC and BIC, while penalizing additional terms.

A. Biological Networks

Biological networks produce a diverse range of functional activities. Regulatory and metabolic networks are critical for cellular function. Breakdown of the function and control circuits of such networks can lead to cancer and other deadly diseases, motivating therapeutic gene modulation or pharmaceutical intervention [6], [7]. Understanding the network structure and the time-dependent dynamic behavior of the signaling molecules can potentially improve our ability to design such interventions [8], [9]. In addition, better models for metabolic networks could facilitate metabolic engineering of designer drugs using microorganisms as bio-chemical factories [30].

Regulatory and metabolic networks have a number of features that make them candidates for dynamic inference using sparsity promoting methods. First these networks are known to be intrinsically sparse, meaning not every component in the network interacts with every other component [31]–[34]. Second, there is growing availability of high-quality experimental data [35], [36]. Third, the system interactions and dynamics have been successfully described using ordinary differential equation models with mass-action kinetics– usually polynomial interactions of order 5 or less. Using these models, systems biologists have characterized a relatively small number of motifs that determine network behavior such as autoregulation, feed-forward loops, feedback loops, and cascades [37]. These network motifs have been detected across bacteria, yeast, plants and animals, suggesting that many biological networks can be characterized in this way [38]. Developing a mechanistic intuition and validating these models usually involves painstaking iteration between theory and experiment [39]. Therefore, we capitalize on the inherent sparsity of the networks and to identify dynamic, mass-action kinetic models, that can be easily interpreted using familiar network motifs.

B. Biological Network Inference

Network inference methods designed for genomic, transcriptomic, proteomic, and metabolomic data include a broad variety of approaches including machine learning, compressive sensing, genetic programing, feature selection, regression, mutual information, correlation, system identification and statistical inference methods (see [1]–[4], [40]–[42] and references therein). Recent methods have been successful in inferring the networked dynamical equations with heuristically selected nonlinear functional forms including: sparse Bayesian inference on a library of Hill functions [42], [43], Bayesian inference on the chemical reaction graph and kinetic parameters using a library of Hill functions [44], [45], an alternating regression using the compact S-system representation [2], [46], and a method combining diffusion maps for dimensionality reduction, a library of Hill functions, and a Markov chain Monte Carlo method for parameter estimation [47]. Alternative methods such as genetic programing and trees [2] can generate very general functional forms, but require significantly more computational time, and risk not comprehensively evaluating model space. In all cases, inferring the connectivity structure of large networks remains an open and challenging problem, which our method circumvents by a fundamentally different construction.

C. Contribution of This Work

In this work, we develop an implicit algorithm to discover networked dynamical systems with rational function nonlinearities. Nonlinear dynamics of metabolic and regulatory networks often include rational terms, motivating this innovation. It is difficult to construct a library containing rational nonlinearities for use with the sparse regression, since a generic rational function is not simply a sparse linear combination of a small set of basis functions. Instead, we write the system as an implicit differential equation in terms of the state and derivatives, and then search for the sparsest vector in the null space of all mixed state and derivative terms, which results in a non-convex optimization problem. To select the model terms in implicit-SINDy, we make use of an optimization formulation by Wright *et al.* [29], and a new non-convex algorithm using the alternating directions method [48], to find the sparsest vector in the null space. This provides an entirely new algorithm to select active rational nonlinearities in the dynamics, which is not possible by simply applying sparse regression in an augmented library. We demonstrate the algorithm to be robust, accurate and fast when applied to three canonical models of biological networks: Michaelis-Menten enzyme kinetics, the regulatory network for competence in bacteria, and the metabolic network for yeast glycolysis.

In the following sections, we provide background on the existing SINDy method [49], and describe the new non-convex sparse selection algorithm. We next validate the algorithm on simulated data for three important biological models: the most fundamental model for enzyme kinetics, a canonical model for regulation of bacterial competence, and a seven-node metabolic network describing glycolysis. Finally we discuss the practical application of implicit-SINDy to real biological systems, including overcoming challenges such as noise and increased system size and incorporating perturbative measurements.

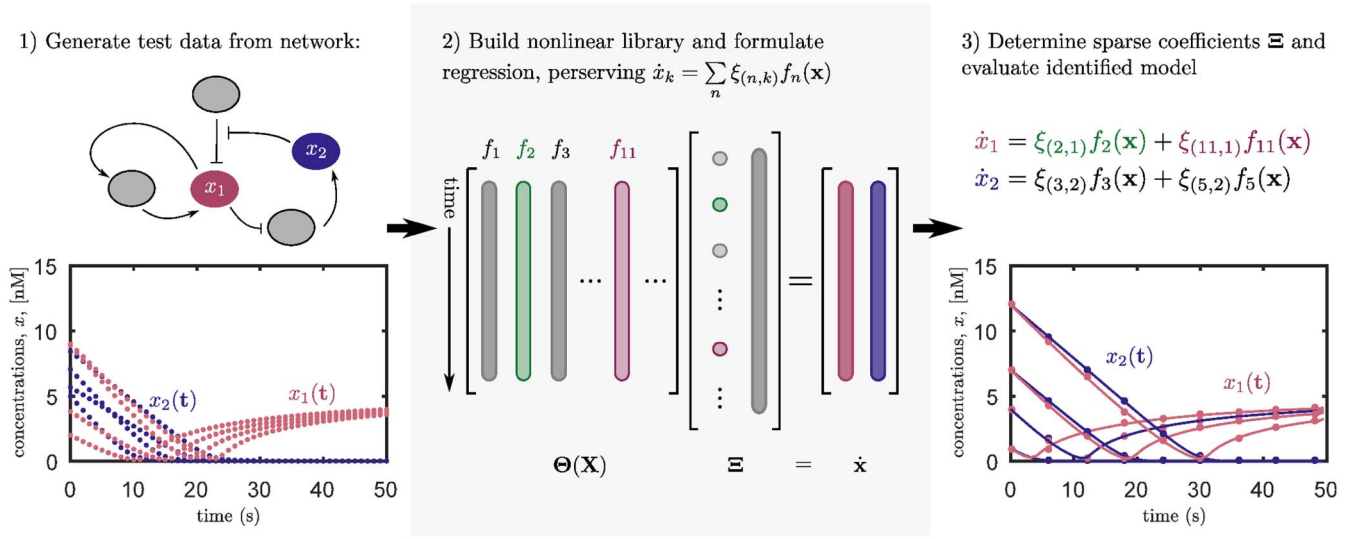

Fig. 1. Methodology for sparse identification of nonlinear dynamics (SINDy) from data. First, data is generated from a dynamical system, in this case a biological network. The time series of data is synthesized into a nonlinear function library, and the terms in this library are related to the time derivative by an overdetermined linear regression problem. Enforcing sparsity ensures that only a small number of coefficients are nonzero, identifying the few active terms in the dynamics that are needed to model the system.

II. Background: Sparse Identification of Nonlinear Dynamics (SINDy)

Discovering dynamical systems from data is an age old pursuit in mathematical physics. Historically, this process relied on a combination of high-quality measurements and expert intuition. With growing computational power and vast quantities of data, the *automated* discovery of dynamical systems and governing equations is a relatively recent phenomenon. Broadly speaking, these techniques may be classified as *system identification*, where methods from statistics and machine learning are used to identify dynamical systems from data.

Nearly all methods of system identification involve some form of regression of data onto dynamics, and the main distinction between the various techniques is the degree to which this regression is constrained. For example, the dynamic mode decomposition (DMD) [50] generates best-fit linear models. Recent nonlinear regression techniques have produced nonlinear dynamic models that preserve physical constraints, such as conservation of energy [51]. Genetic programming has also been used to discover dynamical systems and conservation laws from data [52], [53]. These methods are highly flexible and impose very few constraints on the form of the dynamics identified. More broadly, there is a considerable body of work in the control literature on nonlinear system identification with explicit connections to AIC and BIC [54]–[56].

Here we review the recent sparse identification of nonlinear dynamics (SINDy) method, which leverages advances in machine learning and sparse regression to discover nonlinear dynamical systems from data [49]. SINDy uses sparse regression [57] for improved numerical robustness in noisy overdetermined problems, as opposed to earlier methods [58] that use compressed sensing [59]–[61]. It should be noted that the SINDy method is incapable of handling the rational function nonlinearities present in many biological networks. It is due to this fundamental limitation that we have developed the present implicit-SINDy method.

SINDy relies on the fact that many dynamical systems

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}) \tag{1}$$

are sparse in a given function space. The relevant terms that are active in the dynamics are solved for using an ℓ_1-regularized regression that penalizes the number of active terms. The general framework for SINDy is shown in Fig. 1.

Algorithmically, time-series data is collected from Eq. (1), resulting in a data matrix:

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}(t_1) & \mathbf{x}(t_2) & \cdots & \mathbf{x}(t_m) \end{bmatrix}^T, \tag{2}$$

where ^{T} denotes the matrix transpose. The matrix $\mathbf{X}$ is $m \times n$, where n is the dimension of the state $\mathbf{x} \in \mathbb{R}^n$ and m is the number of measurements of the state in time. For our purposes the state variables are the measured biological components in the network (enzymes, metabolites, transcription factors etc.). Similarly, the matrix of derivatives

$$\dot{\mathbf{X}} = \begin{bmatrix} \dot{\mathbf{x}}(t_1) & \dot{\mathbf{x}}(t_2) & \cdots & \dot{\mathbf{x}}(t_m) \end{bmatrix}^T, \tag{3}$$

is collected or computed from the state data in $\mathbf{X}$; the total-variation regularized derivative [62] provides a numerically robust method to compute derivatives from noisy data.

Next, a library of candidate nonlinear functions, described as features in [63], is constructed from $\mathbf{X}$:

$$\Theta(\mathbf{X}) = \begin{bmatrix} \mathbf{1} & \mathbf{X} & \mathbf{X}^2 & \cdots & \mathbf{X}^d & \cdots & \sin(\mathbf{X}) & \cdots \end{bmatrix}, \tag{4}$$

where $\mathbf{X}^d$ denotes the matrix containing all possible column vectors obtained from time-series of the d-th degree polynomials in the state vector $\mathbf{x}$. For example, for a system with two states $\mathbf{x} = [x_1, x_2]^T$, the matrix $\mathbf{X}^2 = [x_1^2(\mathbf{t}), (x_1 x_2)(\mathbf{t}), x_2^2(\mathbf{t})]$, where $\mathbf{t}$ is a vector of times at which the state is measured. Thus, the vector $\mathbf{x}$ is a symbolic *variable*, while the matrix $\mathbf{X}$ is a *data* matrix.

It is now possible to relate the time derivatives in $\dot{\mathbf{X}}$ to the candidate nonlinearities in $\Theta(\mathbf{X})$ by:

$$\dot{\mathbf{X}} = \Theta(\mathbf{X})\Xi, \tag{5}$$

where each column $\boldsymbol{\xi}_k$ in $\boldsymbol{\Xi}$ is a vector of coefficients that determines which terms are active in the k-th row equation of Eq. (1). To enforce sparsity in the dynamics, we solve for each column $\boldsymbol{\xi}_k$ using sparse regression. One option is to use the LASSO [57]:

$$\boldsymbol{\xi}_k = \text{argmin}_{\boldsymbol{\xi}'_k} \left\| \dot{\mathbf{X}}_k - \boldsymbol{\Theta}(\mathbf{X})\boldsymbol{\xi}'_k \right\|_2 + \lambda \left\| \boldsymbol{\xi}'_k \right\|_1, \tag{6}$$

where $\dot{\mathbf{X}}_k$ is the k-th column of $\dot{\mathbf{X}}$. However, in the original SINDy method, a sequential thresholded least-squares algorithm was implemented instead of LASSO. Once the sparse coefficient vectors $\boldsymbol{\xi}_k$ are determined, a model of the nonlinear dynamical system may be constructed:

$$\dot{x}_k = \boldsymbol{\Theta}(\mathbf{x})\boldsymbol{\xi}_k, \tag{7}$$

where x_k is the kth element of $\mathbf{x}$ and $\boldsymbol{\Theta}(\mathbf{x})$ refers to a row vector whose elements are symbolic functions of $\mathbf{x}$, as opposed to the data matrix $\boldsymbol{\Theta}(\mathbf{X})$.

Using sparse regression to identify active terms in the dynamics from the candidate library $\boldsymbol{\Theta}(\mathbf{X})$ is a convex optimization. The alternative is to apply a separate constrained regression on every possible subset of nonlinearities, and then to choose the model that is both accurate and sparse. This brute-force search is intractable, and the SINDy method makes it possible to select the sparse model in this combinatorially large set of candidate models.

The polynomial and trigonometric nonlinearities in Eq. (4) are sufficient for a large class of dynamical systems. For example, evaluating all polynomials up to order n is equivalent to assuming that the biological network has dynamics determined by mass action kinetics up to n-mers (monomers, dimers, trimers, etc.). However, if there are time-scale separations in the mass action kinetics, fast reactions are effectively at steady state, and the remaining equations contain rational functions [64]. As we consider systems where the dynamics include rational functions, constructing a comprehensive library becomes more complicated. If we generate all rational nonlinearities:

$$f(\mathbf{x}) = \frac{f_N(\mathbf{x})}{f_D(\mathbf{x})}, \tag{8}$$

where $f_N(\mathbf{x})$ and $f_D(\mathbf{x})$ are both polynomial functions, the library would be prohibitively large for computational purposes. Therefore, we develop a computationally tractable framework in the next section for functional library construction that accounts for dynamics with rational functions.

III. Inferring Nonlinear Dynamical Systems With Rational Functions

Many relevant dynamical systems contain rational functions in the dynamics, motivating the need to generalize the SINDy algorithm to include more general nonlinearities than simple polynomial or trigonometric functions. The original SINDy algorithm bypasses the computation and evaluation of all N_p candidate regression models, as enumerated in Section I, by performing a sparse approximation of the dynamics in a library constructed from the N_m candidate monomial features. However, it is not possible to simply apply the original SINDy procedure and include rational functions, since generic rational nonlinearities are not sparse linear combinations of a small number of rational functions. Instead, it is necessary to modify the sparse dynamic regression problem to solve for the sparsest *implicit* ordinary differential equation according to the following procedure.

Consider a dynamical system of the form in Eq. (1), but where the dynamics of each $k = 1, 2, \ldots, n$ variables may contain rational functions:

$$\dot{x}_k = \frac{f_N(\mathbf{x})}{f_D(\mathbf{x})} \tag{9}$$

where $f_N(\mathbf{x})$ and $f_D(\mathbf{x})$ represent numerator and denominator polynomials in the state variable $\mathbf{x}$. It is important to note that these rational terms in an ordinary differential equation are fundamentally *nonlinear*, as opposed to the rational functions used to describe the transfer function of a linear system in the frequency domain. Linear systems theory has a rich history estimating rational transfer functions [65]–[67]. In contrast, (9) describes a nonlinear dynamical system with a rational function in the time domain. For each equation in this nonlinear dynamical system, it is possible to multiply both sides by the denominator polynomial, resulting in the equation:

$$f_N(\mathbf{x}) - f_D(\mathbf{x})\dot{x}_k = 0. \tag{10}$$

The implicit form of Eq. (10) motivates a generalization of the function library $\boldsymbol{\Theta}$ in Eq. (4) in terms of the state $\mathbf{x}$ and the derivative $\dot{x}_k$:

$$\boldsymbol{\Theta}(\mathbf{X}, \dot{x}_k(\mathbf{t})) = \begin{bmatrix} \boldsymbol{\Theta}_N(\mathbf{X}) & \text{diag}(\dot{x}_k(\mathbf{t}))\boldsymbol{\Theta}_D(\mathbf{X}) \end{bmatrix}. \tag{11}$$

The first term, $\boldsymbol{\Theta}_N(\mathbf{X})$, is the library of numerator monomials in $\mathbf{x}$, as in Eq. (4). The second term, $\text{diag}(\dot{x}_k(\mathbf{t}))\boldsymbol{\Theta}_D(\mathbf{X})$, is obtained by multiplying each column of the library of denominator polynomials $\boldsymbol{\Theta}_D(\mathbf{X})$ with the vector $\dot{x}_k(\mathbf{t})$ in an element-wise fashion. For a single variable x_k, this would give the following:

$$\text{diag}(\dot{x}_k(\mathbf{t}))\boldsymbol{\Theta}(\mathbf{X}) = \begin{bmatrix} \dot{x}_k(\mathbf{t}) & (\dot{x}_k x_k)(\mathbf{t}) & \left(\dot{x}_k x_k^2\right)(\mathbf{t}) & \ldots \end{bmatrix}. \tag{12}$$

A schematic of this library is shown in Fig. 2. In most cases, we will use the same polynomial degree for both the numerator and denominator library, so that $\boldsymbol{\Theta}_N(\mathbf{X}) = \boldsymbol{\Theta}_D(\mathbf{X})$. Thus, the augmented library in Eq. (11) is only twice the size of the original polynomial library in Eq. (4).

We may now write the dynamics in Eq. (10) in terms of the augmented library in Eq. (11):

$$\boldsymbol{\Theta}(\mathbf{X}, \dot{x}_k(\mathbf{t}))\boldsymbol{\xi}_k = \mathbf{0}. \tag{13}$$

The sparse vector of coefficients $\boldsymbol{\xi}_k$ will have non-zero entries for the terms active in the nonlinear dynamics. However, it is not possible to use the same method of sparse regression as in the original SINDy algorithm, i.e., to find the sparsest vector $\boldsymbol{\xi}_k$ that satisfies Eq. (13), since the sparsest vector would be identically zero.

To find the sparsest non-zero vector $\boldsymbol{\xi}_k$ that satisfies Eq. (13), we note that any such vector will be in the null space of $\boldsymbol{\Theta}$. After identifying the null space of $\boldsymbol{\Theta}$, we need only find the sparsest vector in this subspace. Although this is a non-convex problem, there are straightforward algorithms based

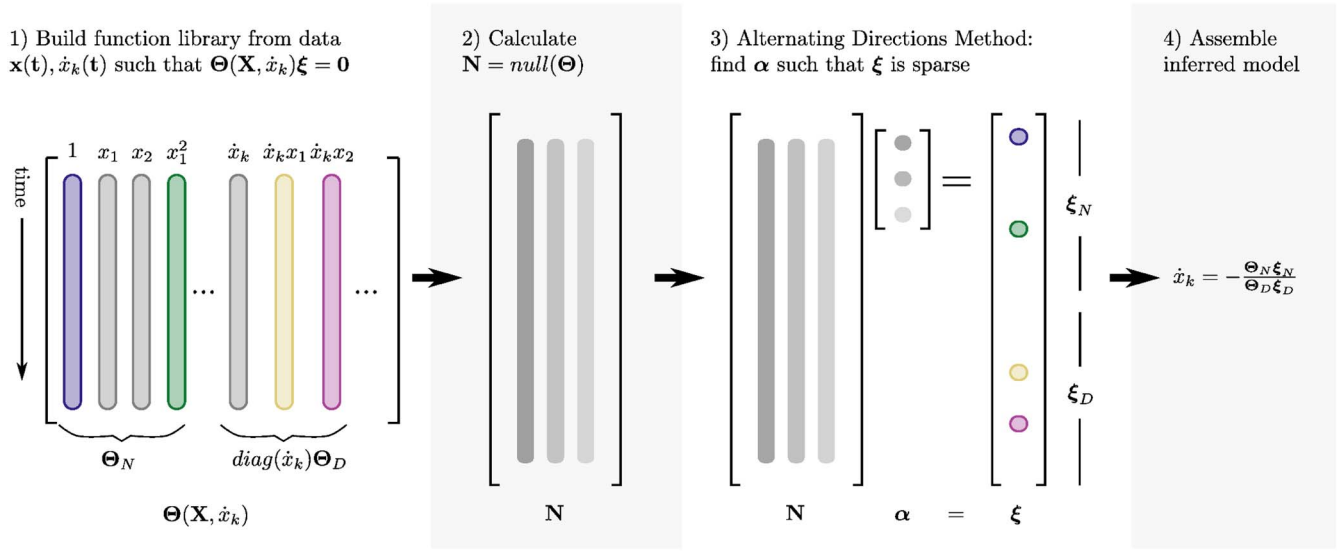

Fig. 2. SINDy algorithm for rational functions. Assemble a matrix, $\mathbf{\Theta}(\mathbf{X}, \dot{x}_k)$ where each column is a nonlinear function evaluated for time series data x_1, x_2, x_3.. and one species derivative $\dot{x}_k(\mathbf{t})$. Next, calculate $\mathbf{N}$ a orthonormal basis for the null space of $\mathbf{\Theta}$. Then, use an alternating directions method [48] to find a sparse vector, $\boldsymbol{\xi}$, in the null space. The sparse vector $\boldsymbol{\xi}$, then satisfies $\mathbf{\Theta}\boldsymbol{\xi} = 0$. Using the sparse coefficients from $\boldsymbol{\xi}$ and the functional library $\mathbf{\Theta}$ assemble the inferred model. This algorithm must be performed for the derivative $\dot{x}_k$ of each species.

on the alternating directions method (ADM) developed by Qu *et al.* [48] to identify the sparsest vector in a subspace.

A. Algorithm for Sparse Selection of Rational Functions

The algorithm for finding $\boldsymbol{\xi}_k$ is illustrated in Fig. 2. First, we build our functional library $\mathbf{\Theta}(\mathbf{X}, \dot{x}_k(\mathbf{t}))$ using both the time series data of the state variables and derivative, as discussed above. Second, we calculate a matrix, $\mathbf{N}$, with columns spanning the null space of $\mathbf{\Theta}$. We wish to find the linear combinations of columns in $\mathbf{N}$ that produces a sparse vector $\boldsymbol{\xi}$. For this third step, we use the alternating directions method developed by Qu *et al.* [48] that finds the sparsest vector in a subspace. We enforce some magnitude of sparsity using a threshold, λ. For the fourth and final step, we select the active nonlinear functions using $\boldsymbol{\xi}$ and $\mathbf{\Theta}$, and assemble the inferred model.

As the appropriate λ is unknown *a priori*, we repeat the third and fourth steps for varying λ. Increasing λ increases the sparsity (decreasing the number of terms) in $\boldsymbol{\xi}$, as shown in Fig. 3A. Each $\boldsymbol{\xi}(\lambda)$ produces an inferred model of varying accuracy and sparsity. From these models we calculate a Pareto front and select the most parsimonious model, as shown in Fig. 3B. A Pareto front is calculated by plotting the number of terms on the x-axis vs an error indicating how well $\boldsymbol{\xi}$ satisfies our implicit equation, $|\mathbf{\Theta}\boldsymbol{\xi}|$, on the y-axis. The most parsimonious model is readily identifiable at the sharp drop-off in error. As we will show, this method succeeds at identifying the correct rational terms and coefficients.

B. General Formulation for Implicit ODEs

The procedure above may be applied to identify more general implicit ordinary differential equations, beyond those just containing rational function nonlinearities. The library $\mathbf{\Theta}(\mathbf{X}, \dot{x}_k(\mathbf{t}))$ contains a subset of the columns of the library

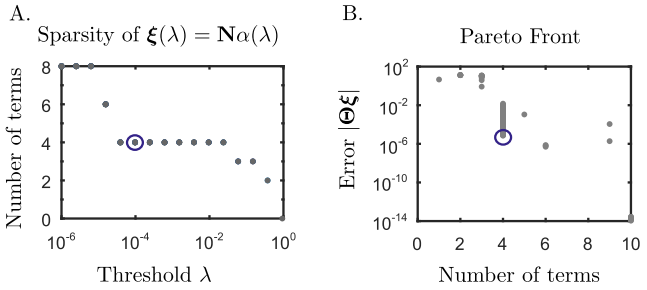

Fig. 3. A. Increasing the sparsity threshold λ during ADM creates coefficient vectors, $\boldsymbol{\xi}$, with monotonically decreasing number of terms. B. For each $\boldsymbol{\xi}(\lambda)$ we calculate an error as $|\mathbf{\Theta}\boldsymbol{\xi}|$, and produce the Pareto Front. For the cases tested here, a large cliff in the error indicates the best choice of $\boldsymbol{\xi}(\lambda)$ (circled on A. and B.) and the most parsimonious model.

$\mathbf{\Theta}([\mathbf{X}\ \ \dot{\mathbf{X}}])$, which is obtained by building nonlinear functions of the state $\mathbf{x}$ and derivative $\dot{\mathbf{x}}$. Identifying the sparsest vector in the null space of $\mathbf{\Theta}([\mathbf{X}\ \ \dot{\mathbf{X}}])$ provides more flexibility in identifying nonlinear equations with mixed terms containing various powers of any combination of derivatives and states. For example, the system given by

$$\dot{x}^3 x - \dot{x}x^2 - x^3 = 0 \tag{14}$$

may be encoded as a sparse vector in the null space of $\mathbf{\Theta}([\mathbf{X}\ \ \dot{\mathbf{X}}])$. It is also straightforward to extend the formulation to include higher order derivatives, by increasing the features in the $\mathbf{\Theta}$ library. For example, second-order implicit dynamical systems may be formulated in the following library:

$$\mathbf{\Theta}([\mathbf{X}\quad \dot{\mathbf{X}}\quad \ddot{\mathbf{X}}]). \tag{15}$$

The generality of this approach enables the identification of many more systems of interest, in addition to those systems with rational function nonlinearities explored below.

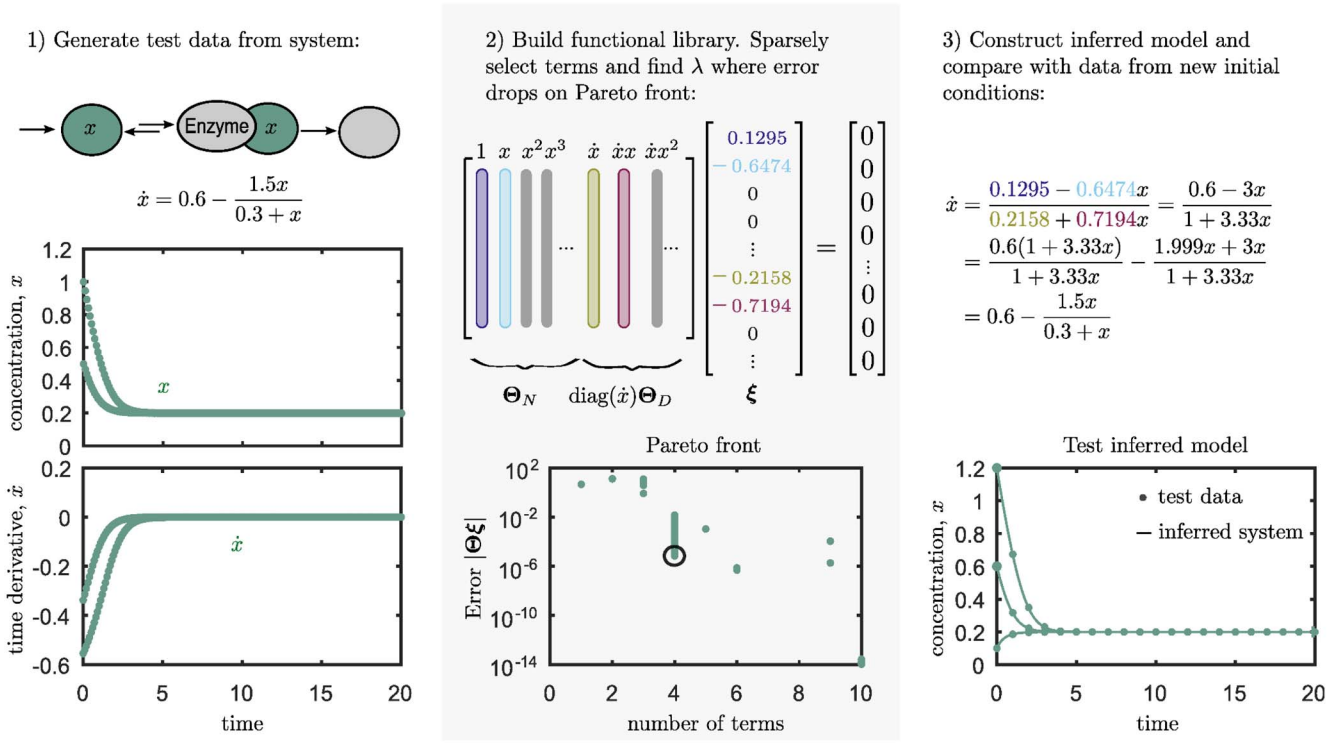

Fig. 4. Algorithm applied to the Michaelis-Menten kinetics for an enzymatic reaction. Step 1) Generate two time series of the single state variable, $x(\mathbf{t})$, and time derivative, $\dot{x}(\mathbf{t})$. Step 2) Discovered active functions and their corresponding coefficients are indicated by color. The error drops sharply at 4 terms on the Pareto front (circled) . The most parsimonious model has four active functions: two in the numerator and two in the denominator (indicated by color). Step 3) Allowing for rational function factorization, the inferred model is equivalent to the original model.

IV. Results

The implicit-SINDy architecture is tested on a number of canonical models of biological networked dynamical systems. Validation of the method on these models allows for potential broader application. We demonstrate that the method is fast, accurate and robust for inferring Michaelis-Menten enzyme kinetics, the regulatory network for competence in bacteria, and the metabolic network for yeast glycolysis.

A. Simple Example: Michaelis-Menten Kinetics

Perhaps the most well known model for enzyme kinetics is the Michaelis-Menten model [68], [69]. This model captures the dynamics of an enzyme binding and unbinding with a substrate (x), and then reacting irreversibly to produce a product, as shown in Fig. 4. A separation of time-scales argument, where binding and unbinding dynamics are fast, or a more general steady state assumption [70], reduces the dynamics to a single state-variable equation with a rational function in the dynamics. Traditionally, biochemists vary the initial concentration of x in a titration experiment to fit the Michaelis-Menten equation to the data.

Using time series data from only two initial concentrations, our algorithm extracts the correct functional form from a larger subset of possible functions and fits the coefficients accurately (Fig. 4). First we generate data from the single dynamic equation

$$\dot{x} = j_x - \frac{V_{max}x}{K_m + x}, \tag{16}$$

with some flux source of x, j_x, and an enzymatic reaction of the Michaelis-Menten form consuming x. Here, V_{max} is the maximum rate of the reaction and K_m is the concentration of half-maximal reaction rate. Generally the time series data of the concentration, $x(\mathbf{t})$, is measurable, while the time series data for the derivative can be calculated from $x(\mathbf{t})$.

Next, we apply implicit-SINDy to determine the coefficient vector $\boldsymbol{\xi}$ and sparsely select the active functions in the dynamics. The library contains polynomial terms up to degree four and has 10 columns. The Pareto front for this system has a sharp drop off in error from around 0.01 to 10^{-5} at four terms, indicating the λ for the most parsimonious model. The associated $\boldsymbol{\xi}$ selects 4 active terms from the function library.

Finally, the rational function constructed from $\boldsymbol{\Theta}$ and $\boldsymbol{\xi}$ needs to be factored to be interpreted as the source flux and Michaelis-Menten terms. When rearranged, the coefficients match the original system. Unsurprisingly, the inferred model matches the original model for time series generated from new initial conditions that were not used in the training data.

B. Regulatory Network: B. Subtilis Competence

Having shown that our method works for the simplest rational model relevant to biological networks, we next test it on a regulatory model with two state variables [71]. Süel *et al.* [71] demonstrated that a dynamic gene network enables cells to switch between multiple behaviors – in this case *B. subtilis* bacteria switch between taking up DNA from the environment (competence) and vegetative growth. Other regulatory

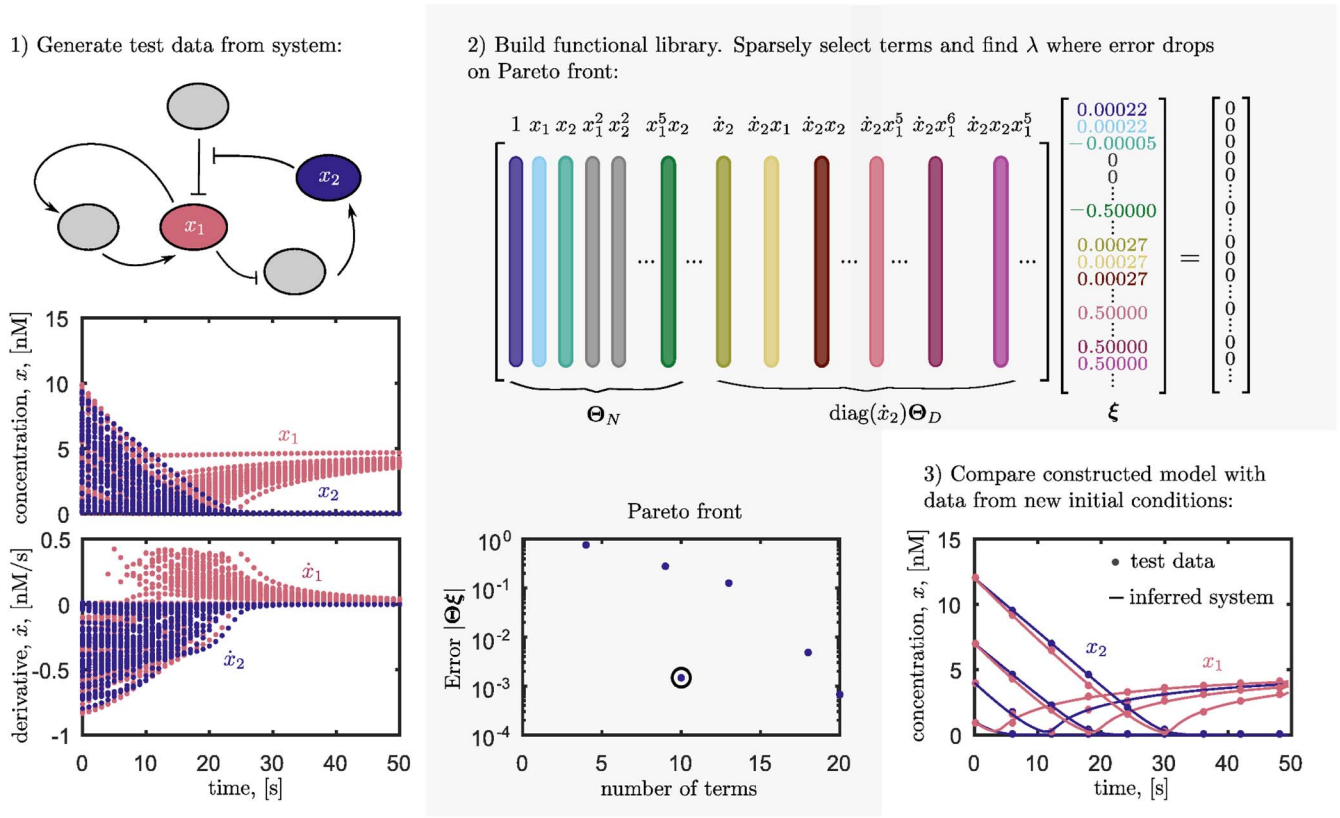

Fig. 5. Algorithm applied to a regulatory network with two measured state-variables: x_1 (blue) and x_2 (pink). Pointed arrows indicate activation and blunted arrows indicate repression. The functional library $\boldsymbol{\Theta}$, sparse vector $\boldsymbol{\xi}$, and Pareto Front for x_2 are shown in Step 2. 10 of 56 terms are active in the library for the most parsimonious model: 4 in the numerator and 6 in the denominator. The inferred model is nearly equivalent to the original model for test data.

networks such as the circadian clock [72], [73] and cell cycle oscillators have been successfully described using models with similar structure and dynamics. In particular, similar dynamics may drive cancer-relevant systems like the tumor suppressor p53 [74].

The dynamics of bacterial competence regulatory system with two states can be described by the following two non-dimensional equations:

$$\dot{x}_1 = a_1 + \frac{a_2 x_1^2}{a_3 + x_1^2} - \frac{x_1}{1 + x_1 + x_2}, \tag{17a}$$

$$\dot{x}_2 = \frac{b_1}{1 + b_2 x_1^5} - \frac{x_2}{1 + x_1 + x_2}. \tag{17b}$$

These two equations are a reduction of dynamical system with six states. Each rational function arises from a steady state (or time-scale separation) assumption about the regulatory processes. The second term (scaled by a_2) in Eq. (17a) represents protein x_1, ComK, activating its own production in an autoregulatory, positive feedback loop. The first term (scaled by b_1) in Eq. (17b), describes x_1 mediated repression of x_2, ComS, in a negative feedback loop. Both of these terms have a Hill-function form, where the power indicates the number of x_1 proteins involved cooperatively in the regulatory complex [37]. The combination of positive and negative feedback results in the network's functional capabilities. The last term in Eqs. (17a) and (17b) describes degradation of x_1 and x_2, mediated by a third unmeasured protein, MecA.

Using this model we generate 40 time series for the regulatory system, as shown in Fig. 5. This model challenges our method in two ways. First, the method must correctly identify the dynamic dependence on two state variables. Second, the model contains polynomial functions up to the 5th degree in the denominator of one term in Eq. (17b). To include this term, the library must contain polynomials up to degree six. Even without knowing the highest polynomial power ahead of time, it is possible to use the implicit-SINDy by trying libraries of increasing polynomial degree. If the library does not have all the required terms, there will be no clear drop off in the Pareto front as there is in Fig. 5.

The library with degree six polynomials in two state variables contains 56 columns, of which 10 are active in the most parsimonious model for x_2 dynamics. The constructed models for x_1 and x_2 match almost exactly with test data generated from the original model. As with our first example, the extracted rational function can be factored to recover exactly the form of Eq. (17b). Additionally, the coefficients identified are within 2% error of the true coefficients shown in Table I.

C. Metabolic Network: Yeast Glycolysis

As a final example, we test our method on a metabolic network. Glycolysis, the process of breaking down glucose to extract energy (ATP and NADPH), is part of central metabolism for all cells. Uncovering the metabolic network for glycolysis took over 100 years from its initial discovery

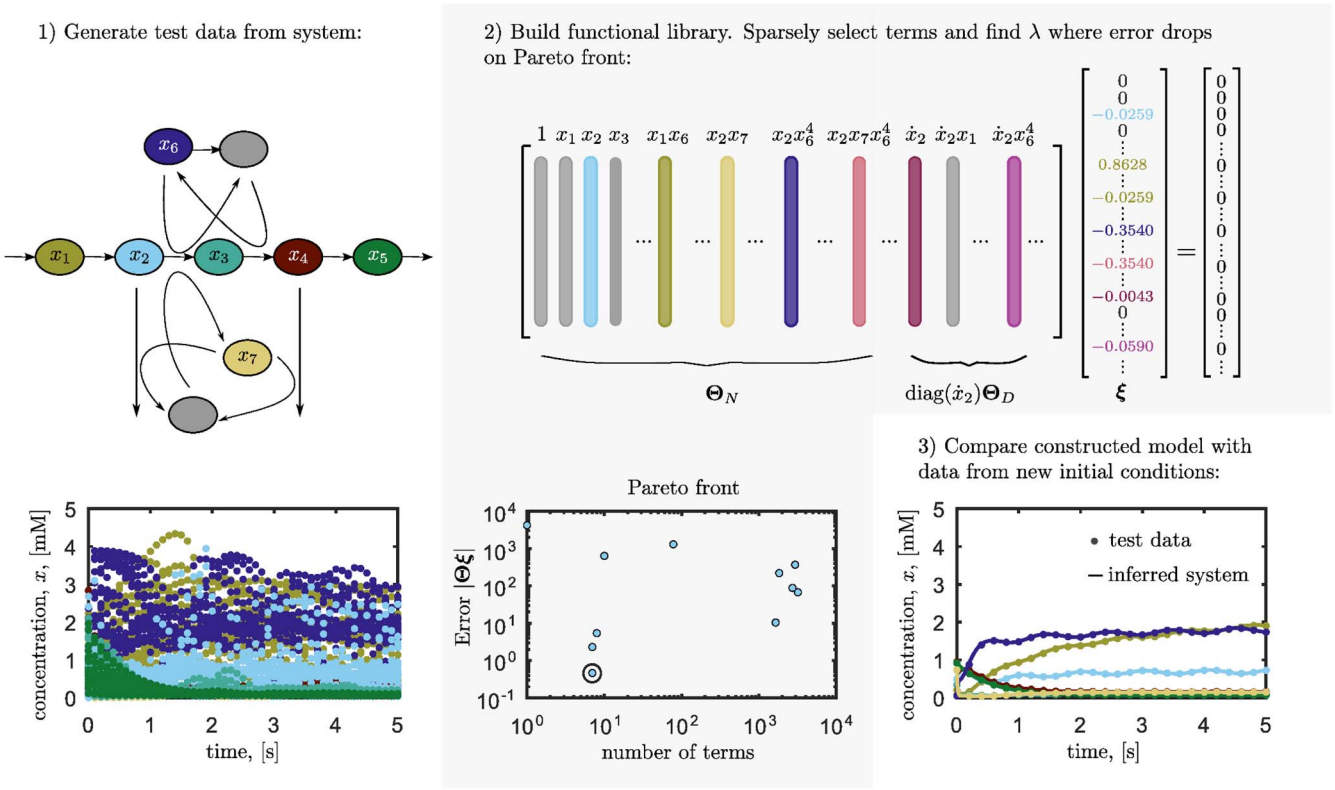

Fig. 6. Algorithm applied to a metabolic network for yeast glycolysis. Step 1). The 7 measured state variables are indicated with separate colors consistent with time series. Arrows indicate reactions between components. Not all data required to infer the network is plotted. Step 2) shows functional library $\boldsymbol{\Theta}$ and corresponding coefficient vector $\boldsymbol{\xi}$ for x_2. Seven functions are active: 5 in the numerator and 2 in the denominator. Step 3) The inferred model is nearly equivalent to the original system for test data. One time series is shown for all 7 state variables.

TABLE I
PARAMETER IDENTIFICATION FOR REGULATORY NETWORK

Parameter	units	True Value	Extracted value
a_1	[nM/s]	0.004	0.00393
a_2	[nM/s]	0.07	0.07006
a_3	[nM]	0.04	0.04000
b_1	[nM/s]	0.82	0.8148
b_2	[nM]	1854.5	1851.9

by Pasteur [75]. Accelerated inference of metabolic networks would aid metabolic disease intervention [76]. Bacteria perform a wide range of yet-to-be discovered chemistry which could be harnessed through metabolic engineering to produce high-value products such as drugs and biofuels [77], [78].

Not only does the yeast glycolysis model we analyze have a larger number of interacting state variables, but it is oscillatory [79]. This network has also been previously analyzed as a test case for model inference [80]. The network shown in Fig. 6 has three equations with rational functions and four with polynomials:

$$\dot{x}_1 = c_1 + \frac{c_2 x_1 x_6}{1 + c_3 x_6^4} \tag{18a}$$

$$\dot{x}_2 = \frac{d_1 x_1 x_6}{1 + d_2 x_6^4} + d_3 x_2 - d_4 x_2 x_7 \tag{18b}$$

$$\dot{x}_3 = e_1 x_2 + e_2 x_3 + e_3 x_2 x_7 + e_4 x_3 x_6 \tag{18c}$$

$$\dot{x}_4 = f_1 x_3 + f_2 x_4 + f_3 x_5 + f_4 x_3 x_6 + f_5 x_4 x_7 \tag{18d}$$

$$\dot{x}_5 = g_1 x_4 + g_2 x_5 \tag{18e}$$

$$\dot{x}_6 = \frac{h_1 x_1 x_6}{1 + h_2 x_6^4} + h_3 x_3 + h_5 x_6 + h_4 x_3 x_7 \tag{18f}$$

$$\dot{x}_7 = j_1 x_2 + j_2 x_2 x_7 + j_3 x_4 x_7. \tag{18g}$$

Given sufficient data, implicit-SINDy correctly infers the network structure and coefficients. In Fig. 6, we show the sparsely selected terms and Pareto front for Eq. (18c). The method correctly selects 7 terms from a library of 3432 functions: 5 in the numerator and 2 in the denominator. Table II shows the true and extracted coefficient values for the model. Some of the parameters, c_1 for example, had incorrect functional dependence on x_6 after factoring the discovered polynomial. However these dependencies were very small, meaning the coefficient of the first erroneous term in a Taylor expansion was $< 0.1\%$ of the magnitude of the correct leading term.

The equations for x_3, x_4, x_5, and x_7 did not require a library with polynomials up to degree six, and could be inferred more quickly. On the other hand, Eq. (18g), required over twice as many measurements as the other equations with rational functions (Eqs. (18c) and (18b)).

D. Noise and Measurements With Implicit-SINDy

The implicit-SINDy algorithm can recover the underlying dynamic equations with the addition of noise. Degraded data compromises the calculation of the null space in Eq. 13 causing all of the singular values of $\boldsymbol{\Theta}$ to be nonzero. Recent work by Gavish and Donoho provides an optimal singular value threshold for truncating the SVD of a data matrix corrupted

TABLE II
PARAMETER IDENTIFICATION FOR METABOLIC NETWORK

Parameter	units	True Value	Extracted value
c_1	[mM/min]	2.5	2.5002*
c_2	[1/(mM min)]	-100	-99.7979
c_3	[$\text{mM}^{-1/4}$]	13.6769	13.6489
d_1	[1/(mM min)]	200	200.6512
d_2	[$\text{mM}^{-1/4}$]	13.6769	13.7209
d_3	[1/min]	-6	-5.9998*
d_4	[1/(mM min)]	-6	-5.9998*
e_1	[1/min]	6	6.0133
e_2	[1/min]	-64	-64.140
e_3	[1/(mM min)]	6	-6.0133
e_4	[1/(mM min)]	16	16.0333
f_1	[1/min]	64	63.6145
f_2	[1/min]	-13	-12.9277
f_3	[1/min]	13	12.9277
f_4	[1/(mM min)]	-16	-15.9036
f_5	[1/(mM min)]	-100	-99.3976
g_1	[1/min]	1.3	1.3002
g_2	[1/min]	-3.1	-3.1003
h_1	[1/(mM min)]	-200	$-200.0^{\dagger}$
h_2	[$\text{mM}^{-1/4}$]	13.6769	$13.6769^{\dagger}$
h_3	[1/min]	128	$128.0^{\dagger}$
h_4	[1/(mM min)]	-1.28	$-1.280^{\dagger}$
h_5	[1/min]	-32	$-32.00^{\dagger}$
j_1	[1/min]	6	6.0102
j_2	[1/(mM min)]	-18	-18.0408
j_3	[1/(mM min)]	-100	-100.2449

Values with a * have errors with functional dependence on x_6. We show the leading order term assuming a Taylor series expansion. † Eq. (18g) required more data to identify, and this allowed extraction of the coefficients to a higher precision.

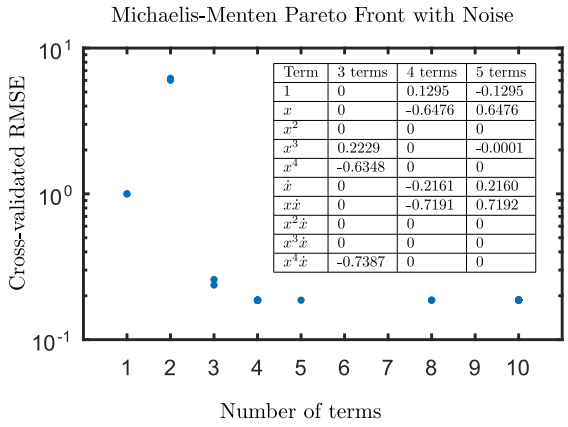

Fig. 7. Pareto front for the Michaelis Menten model from section IV-A with measurement noise added to the concentration time series. Noise is normally distributed with standard deviation $\sigma = 10^{-4}$. There is a plateau in the cross validated root-mean-squared-error (RMSE) at 4 terms. The inset table shows the coefficients for the models corresponding to lowest error for 3, 4, and 5 terms.

with noise [81]. This principled truncation criteria can be efficiently incorporated into the algorithm for implicit-SINDy by approximating the data library $\boldsymbol{\Theta}$ allowing for the null space calculation. As a specific example, we can recover the Michaelis-Menten governing equation in Section IV-A with the addition of normally distributed noise of magnitude $\sigma = 10^{-4}$ (SNR $O(10^4)$). Further, the error of the reconstruction is shown with the Pareto analysis shown in Fig. 7. The error was calculated by reconstructing each inferred model for varying levels of sparsity and cross validating with time series data from

TABLE III
SIZE OF PROBLEM AND DATA REQUIRED TO INFER NETWORK DYNAMICS

# state variables	order of polynomial, d	size library	# time series measurements
1	4	10	2
2	6	56	40
7	6	3432	900 (†2250)

† Eq. (18g) required more data to identify.

three new initial conditions. The table inset in Fig. 7 shows the recovered coefficients for the lowest cross validated error models with 3, 4, and 5 terms. The coefficients for the 4 term model are within 0.05% root mean square error of the coefficients recovered without noise (see Fig. 4). Note, the 5 term model is the same as the 4 term but with an extra, small coefficient that survived the larger threshold value.

Table III presents the amount of data required for each system explored in this article. The measurements are generated by randomizing initial conditions within a range that is reasonable for each natural system. These numerical measurements are intrinsically information-poor, as they often overlap. We anticipate a substantial decrease in the number of measurements required for perturbation experiments, such as knockouts or knockdowns, which have been shown to have much higher information content, see Supplement of Ref. [1, Sec. 5.2].

V. CONCLUSION

In this work we developed an implicitly formulated method for sparse identification of nonlinear dynamics: implicit-SINDy. The method allows for the construction of nonlinear dynamics with rational functions using a library of functional forms that is still computationally manageable for reasonably-sized biological networks. An alternating directions method for selection of a sparse vector in the null space of the library [48] enables us to find and construct a parsimonious model from the full library. Using implicit-SINDy on time-series data computationally generated from three biological models (enzyme kinetics, regulation, and metabolism), we are able to accurately reconstruct the underlying system in each case. Indeed, we correctly recover the coefficients to within 2% of the original values. These results make implicit-SINDy a promising method for model discovery of biological networks.

SINDy is a *data-driven* methodology, meaning it selects the connectivity and dynamics based on the information content of the data alone. It has many advantages, including the fact that there is no parameter tuning in the inferred models aside from a sparsity threshold which is determined by a Pareto front. Moreover, implicit-SINDy greatly expands our ability to rapidly select a model from a large class of candidate dynamical systems, even when nonlinear derivative terms are present. In practice, the method functions much like a highly efficient unsupervised learning algorithm, sparsely selecting dynamics from a large library of possibilities. It differs from information theoretic techniques where a number of viable models are posited and selection of the best model is based upon the minimization of information loss. Such alternative techniques rely

on physical insight (supervised learning) to generate individual models, thus potentially limiting the dynamical systems considered. Given the large number of biological models driven by mass-action kinetics, the implicit-SINDy method can be a critically enabling method for data-driven discovery of underlying biological principals.

There are many intriguing future directions for the method, both in theory and practice. A rigorous connection between the selection process for implict-SINDy and information criteria such as AIC and BIC remains an open question. There are strong similarities in the selection process between information criteria and the Pareto analysis for implicity-SINDy. However, there is an ambiguity in choosing an error metric and log-likelihood interpretation for IC to nonlinear dynamic systems, as discussed in [21]. Application of information criteria to the models discovered by implicit-SINDy would allow quantitative comparison between models. Further, relative-information criteria would provide a systematic strength-of-evidence comparison between models [82]. We expect that these methods could differentiate between true model identification and false drop off on the Pareto front when the functional library is missing terms, an essential state-variable is not included, or the noise level is too high. Further, these methods can be extended to handle large signal-to-noise ratios and stochastic variation, including experimentally relevant data studies such as ensemble averaging and non-regular data acquisition for linear and minimally non-linear regression models [4]. We believe establishing a rigorous connection between information theory metrics and sparse selection will address these challenges for implementation on experimental data.

An additional challenge is in choosing the correct variables with which to build a regression model. Recent work has indicated that even if key dynamic variables are not included in the measurements, it may be possible to obtain regression models using delay coordinates. This is based on the celebrated Takens embedding theorem, which states that under certain conditions, delay coordinates on limited measurements may provide an attractor that is diffeomorphic to the attractor in the original dynamic variables. The use of delay coordinates for dynamic regression is a promising and actively developing field of research [49], [83]. The next critical step in building a dynamic regression model is constructing a library of candidate nonlinearities that is complete enough to capture the system under consideration. In the examples in this paper, dynamics are known to be governed by rational nonlinearities, motivating the particular approach developed. More generally, additional nonlinear terms may be required for other problems. It is also possible to concatenate multiple libraries of candidate nonlinearities. These future directions build upon the success already demonstrated by the implicit-SINDy mathematical architecture.

ACKNOWLEDGMENT

J. L. Proctor and N. M. Mangan 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.

REFERENCES

[1] D. Marbach *et al.*, "Wisdom of crowds for robust gene network inference," *Nat Methods*, vol. 9, no. 8, pp. 796–804, 2012.
[2] T. Cakir and M. J. Khatibipour, "Metabolic network discovery by top-down and bottom-up approaches and paths for reconciliation," *Front. Bioeng. Biotechnol.*, vol. 2, p. 62, Dec. 2014.
[3] W.-X. Wang, Y.-C. Lai, and C. Grebogi, "Data based identification and prediction of nonlinear and complex dynamical systems," *Phys. Rep.*, vol. 644, pp. 1–76, Jul. 2016.
[4] C. J. Oates and S. Mukherjee, "Network inference and biological dynamics," *Ann. Appl. Stat.*, vol. 6, no. 3, pp. 1209–1235, 2012.
[5] D. Chasman, A. F. Siahpirani, and S. Roy, "Network-based approaches for analysis of complex biological systems," *Current Opin. Biotechnol.*, vol. 39, pp. 157–166, Jun. 2016.
[6] R. M. Plenge, E. M. Scolnick, and D. Altshuler, "Validating therapeutic targets through human genetics," *Nat. Rev. Drug Disc.*, vol. 12, no. 8, pp. 581–594, 2013.
[7] H. Yin *et al.*, "Non-viral vectors for gene-based therapy," *Nat. Rev. Genet.*, vol. 15, no. 8, pp. 541–555, 2014.
[8] J. E. Purvis and G. Lahav, "Encoding and decoding cellular information through signaling dynamics," *Cell*, vol. 152, no. 5, pp. 945–956, 2013.
[9] H.-S. Song, F. DeVilbiss, and D. Ramkrishna, "Modeling metabolic systems: The need for dynamics," *Current Opin. Chem. Eng.*, vol. 2, no. 4, pp. 373–382, 2013.
[10] S. Kullback and R. A. Leibler, "On information and sufficiency," *Ann. Math. Stat.*, vol. 22, no. 1, pp. 79–86, Mar. 1951.
[11] S. Kullback, *Information Theory and Statistics*. New York, NY, USA: Wiley, 1959.
[12] H. Akaike, "Information theory and an extension of the maximum likelihood principle," in *2nd International Symposium on Information Theory, Tsahkadsor, Armenia, USSR, September 2-8, 1971*, B. N. Petrov and F. Csáki, Eds. Budapest, Hungary: Akadémiai Kiadó, 1973, pp. 267–281.
[13] H. Akaike, "A new look at the statistical model identification," *IEEE Trans. Autom. Control*, vol. 19, no. 6, pp. 716–723, Dec. 1974.
[14] K. P. Burnham and D. R. Anderson, *Model Selection and Multi-Model Inference*, 2nd ed. New York, NY, USA: Springer, 2002.
[15] L. Kuepfer, M. Peter, U. Sauer, and J. Stelling, "Ensemble modeling for analysis of cell signaling dynamics," *Nat. Biotechnol.*, vol. 25, no. 9, pp. 1001–1006, 2007.
[16] J. Schaber *et al.*, "Automated ensemble modeling with *modelMaGe*: Analyzing feedback mechanisms in the Sho1 branch of the HOG pathway," *PLoS ONE*, vol. 6, no. 3, pp. 1–7, 2011.
[17] M. Woodward, *Epidemiology: Study Design and Data Analysis*, 2nd ed. Boca Raton, FL, USA: Chapman & Hall, 2004.
[18] I. M. Blake *et al.*, "The role of older children and adults in wild poliovirus transmission," *Proc. Nat. Academy Sci.*, vol. 111, no. 29, pp. 10604–10609, 2014.
[19] G. Schwarz, "Estimating the dimension of a model," *Ann. Stat.*, vol. 6, no. 2, pp. 461–464, 1978.
[20] G. Claeskens and N. L. Hjorth, *Model Selection and Model Averaging*. Cambridge, U.K.: Cambridge Univ. Press, 2008.
[21] T. Nakamura, K. Judd, A. I. Mees, and M. Small, "A comparative study of information criteria for model selection," *Int. J. Bifurcations Chaos*, vol. 16, no. 8, pp. 2153–2175, 2006.
[22] W. D. Penny, "Comparing dynamic causal models using AIC, BIC and free energy," *NeuroImage*, vol. 59, no. 1, pp. 319–330, 2012.
[23] A. Tjärnberg, T. E. M. Nordling, M. Studham, and E. L. L. Sonnhammer, "Optimal sparsity criteria for network inference," *J. Comput. Biol.*, vol. 20, no. 5, pp. 398–408, 2013.
[24] F. Büchel *et al.*, "Path2Models: Large-scale generation of computational models from biochemical pathway maps," *BMC Syst. Biol.*, vol. 7, no. 1, p. 116, 2013.
[25] P. R. Cohen, "DARPA's big mechanism program," *Phys. Biol.*, vol. 12, no. 4, 2015, Art. no. 045008.
[26] C. M. Bishop, *Pattern Recognition and Machine Learning*, vol. 1. New York, NY, USA: Springer, 2006.
[27] K. P. Murphy, *Machine Learning: A Probabilistic Perspective*. Cambridge, MA, USA: MIT Press, 2012.
[28] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An Introduction to Statistical Learning*. New York, NY, USA: Springer, 2013.
[29] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 210–227, Feb. 2009.
[30] J. D. Keasling, "Manufacturing molecules through metabolic engineering," *Science*, vol. 330, no. 6009, pp. 1355–1358, 2010.

[31] M. G. Grigorov, "Global properties of biological networks," *Drug Disc. Today*, vol. 10, no. 5, pp. 365–372, 2005.
[32] M. Andrecut and S. A. Kauffman, "On the sparse reconstruction of gene networks," *J. Comput. Biol.*, vol. 15, no. 1, pp. 21–30, 2008.
[33] M. Öksüz, H. Sadikoğlu, and T. Çakir, "Sparsity as cellular objective to infer directed metabolic networks from steady-state metabolome data: A theoretical analysis," *PLoS ONE*, vol. 8, no. 12, pp. 1–7, 2013.
[34] T. E. M. Nordling and E. W. Jacobsen, "On sparsity as a criterion in reconstructing biochemical networks," in *Proc. 18th Int. Federation Autom. Control (IFAC) World Congr.*, Milan, Italy, 2011, pp. 11672–11678.
[35] D. G. Spiller, C. D. Wood, D. A. Rand, and M. R. H. White, "Measurement of single-cell dynamics," *Nature*, vol. 465, pp. 736–745, Jun. 2010.
[36] M. Wu and A. K. Singh, "Single-cell protein analysis," *Current Opin. Biotechnol.*, vol. 23, no. 1, pp. 83–88, 2012.
[37] U. Alon, *An Introduction to Systems Biology: Design Principles of Biological Circuits*. Boca Raton, FL, USA: Chapman & Hall, 2007.
[38] U. Alon, "Network motifs: Theory and experimental approaches," *Nat. Rev. Genet.*, vol. 8, no. 6, pp. 450–461, 2007.
[39] S. L. Spencer and P. K. Sorger, "Measuring and modeling apoptosis in single cells," *Cell*, vol. 144, no. 6, pp. 926–939, 2011.
[40] J. Roll, A. Nazin, and L. Ljung, "Nonlinear system identification via direct weight optimization," *Automatica*, vol. 41, no. 3, pp. 475–490, 2005.
[41] G. Stolovitzky, D. Monroe, and A. Califano, "Dialogue on reverse-engineering assessment and methods: The DREAM of high-throughput pathway inference," *Ann. New York Acad. Sci.*, vol. 1115, no. 1, pp. 1–22, 2007.
[42] W. Pan, Y. Yuan, J. Gonçalves, and G.-B. Stan, "A sparse Bayesian approach to the identification of nonlinear state-space systems," *IEEE Trans. Autom. Control*, vol. 61, no. 1, pp. 182–187, Jan. 2016.
[43] W. Pan, Y. Yuan, J. Gonçalves, and G.-B. Stan, "Reconstruction of arbitrary biochemical reaction networks: A compressive sensing approach," in *Proc. IEEE Conf. Decis. Control*, Maui, HI, USA, 2012, pp. 2334–2339.
[44] C. J. Oates *et al.*, "Causal network inference using biochemical kinetics," *Bioinformatics*, vol. 30, no. 17, pp. I468–I474, 2014.
[45] K. A. McGoff *et al.*, "The local edge machine: Inference of dynamic models of gene regulation," *Genome Biol.*, vol. 17, no. 1, p. 214, 2016.
[46] B. C. Daniels and I. Nemenman, "Efficient inference of parsimonious phenomenological models of cellular dynamics using S-systems and alternating regression," *PLoS ONE*, vol. 10, no. 3, 2015, Art. no. e0119821.
[47] A. Ocone, L. Haghverdi, N. S. Mueller, and F. J. Theis, "Reconstructing gene regulatory dynamics from high-dimensional single-cell snapshot data," *Bioinformatics*, vol. 31, no. 12, pp. I89–I96, 2015.
[48] Q. Qu, J. Sun, and J. Wright, "Finding a sparse vector in a subspace: Linear sparsity using alternating directions," in *Proc. Adv. Neural Inf. Process. Syst.*, Montreal, QC, USA, 2014, pp. 3401–3409.
[49] S. L. Brunton, J. L. Proctor, and J. N. Kutz, "Discovering governing equations from data by sparse identification of nonlinear dynamical systems," *Proc. Nat. Acad. Sci.*, vol. 113, no. 15, pp. 3932–3937, 2016.
[50] J. N. Kutz, S. L. Brunton, B. W. Brunton, and J. L. Proctor, *Dynamic Mode Decomposition: Data-Driven Characterization of Complex Systems*. Philadelphia, PA, USA: SIAM, 2016.
[51] A. J. Majda and J. Harlim, "Physics constrained nonlinear regression models for time series," *Nonlinearity*, vol. 26, no. 1, p. 201, 2012.
[52] J. Bongard and H. Lipson, "Automated reverse engineering of nonlinear dynamical systems," *Proc. Nat. Acad. Sci.*, vol. 104, no. 24, pp. 9943–9948, 2007.
[53] M. Schmidt and H. Lipson, "Distilling free-form natural laws from experimental data," *Science*, vol. 324, no. 5923, pp. 81–85, 2009.
[54] J. Paduart *et al.*, "Identification of nonlinear systems using polynomial nonlinear state space models," *Automatica*, vol. 46, no. 4, pp. 647–656, 2010.
[55] G. Pillonetto, F. Dinuzzo, T. Chen, G. De Nicolao, and L. Ljung, "Kernel methods in system identification, machine learning and function estimation: A survey," *Automatica*, vol. 50, no. 3, pp. 657–682, 2014.
[56] T. Chen, M. S. Andersen, L. Ljung, A. Chiuso, and G. Pillonetto, "System identification via sparse multiple kernel-based regularization using sequential convex optimization techniques," *IEEE Trans. Autom. Control*, vol. 59, no. 11, pp. 2933–2945, Nov. 2014.
[57] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. Roy. Stat. Soc.*, vol. 58, no. 1, pp. 267–288, 1996.
[58] W. X. Wang, R. Yang, Y.-C. Lai, V. Kovanis, and C. Grebogi, "Predicting catastrophes in nonlinear dynamical systems by compressive sensing," *Phys. Rev. Lett.*, vol. 106, no. 15, pp. 1–4, 2011.
[59] E. J. Candès, "Compressive sensing," in *Proc. Int. Congr. Math.*, vol. 3, 2006, pp. 1433–1452.
[60] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.
[61] R. G. Baraniuk, "Compressive sensing," *IEEE Signal Process. Mag.*, vol. 24, no. 4, pp. 118–120, Jul. 2007.
[62] R. Chartrand, "Numerical differentiation of noisy, nonsmooth data," *ISRN Appl. Math.*, vol. 2011, 2011, Art. no. 164564.
[63] I. Arnaldo, U.-M. O'Reilly, and K. Veeramachaneni, "Building predictive models via feature synthesis," in *Proc. Annu. Conf. Genet. Evol. Comput.*, Madrid, Spain, 2015, pp. 983–990.
[64] J. Gunawardena, "Time-scale separation—Michaelis and Menten's old idea, still bearing fruit," *FEBS J.*, vol. 281, no. 2, pp. 473–488, 2014.
[65] R. Kalman, "Design of a self-optimizing control system," *Trans. ASME*, vol. 80, pp. 468–478, 1958.
[66] S. A. Billings and S. Chen, "Identification of non-linear rational systems using a prediction-error estimation algorithm," *Int. J. Syst. Sci.*, vol. 20, no. 3, pp. 467–494, 1989.
[67] S. Chen, S. A. Billings, and P. M. Grant, "Non-linear system identification using neural networks," *Int. J. Control*, vol. 51, no. 6, pp. 1191–1214, 1990.
[68] L. Michaelis and M. L. Menten, "Die kinetik der Invertinwirkung," *Biochem Z*, vol. 49, pp. 333–369, 1913.
[69] K. A. Johnson and R. S. Goody, "The original michaelis constant: Translation of the 1913 Michaelis–Menten paper," *Biochemistry*, vol. 50, no. 39, pp. 8264–8269, 2011.
[70] G. E. Briggs, "A further note on the kinetics of enzyme action," *Biochem. J.*, vol. 19, no. 6, pp. 1037–1038, 1925.
[71] G. M. Süel, J. Garcia-Ojalvo, L. M. Liberman, and M. B. Elowitz, "An excitable gene regulatory circuit induces transient cellular differentiation," *Nature*, vol. 440, no. 7083, pp. 545–550, 2006.
[72] P. E. Hardin, "The circadian timekeeping system of Drosophila," *Current Biol.*, vol. 15, no. 17, pp. R714–R722, 2005.
[73] J.-C. Leloup and A. Goldbeter, "Toward a detailed computational model for the mammalian circadian clock," *Proc. Nat. Acad. Sci.*, vol. 100, no. 12, pp. 7051–7056, 2003.
[74] E. Batchelor, A. Loewer, and G. Lahav, "The ups and downs of p53: Understanding protein dynamics in single cells," *Nature Rev. Cancer*, vol. 9, no. 5, pp. 371–377, 2009.
[75] E. Racker, "From Pasteur to Mitchell: A hundred years of bioenergetics," *Fed. Proc.*, vol. 39, no. 2, pp. 210–215, Feb. 1980.
[76] M. L. Yarmush and S. Banta, "Metabolic engineering: Advances in modeling and intervention in health and disease," *Annu. Rev. Biomed. Eng.*, vol. 5, no. 1, pp. 349–381, 2003.
[77] D. I. Ellis and R. Goodacre, "Metabolomics-assisted synthetic biology," *Current Opinion Biotech.*, vol. 23, no. 1, pp. 22–28, 2012.
[78] D. C. Ducat, J. C. Way, and P. A. Silver, "Engineering cyanobacteria to generate high-value products," *Trends Biotech.*, vol. 29, no. 2, pp. 95–103, Feb. 2011.
[79] J. Wolf and R. Heinrich, "Effect of cellular interaction on glycolytic oscillations in yeast: A theoretical investigation." *Biochem. J.*, vol. 345, no. 2, pp. 321–334, 2000.
[80] M. D. Schmidt *et al.*, "Automated refinement and inference of analytical models for metabolic networks," *Phys. Biol.*, vol. 8, no. 5, 2011, Art. no. 055011.
[81] M. Gavish and D. L. Donoho, "Optimal shrinkage of singular values," *IEEE Trans. Inf. Theory*, vol. 60, no. 8, pp. 5040–5053, 2014.
[82] K. P. Burnham and D. R. Anderson, "Multimodel inference: Understanding AIC and BIC in model selection," *Sociol. Methods Res.*, vol. 33, no. 2, pp. 261–304, 2004.
[83] S. L. Brunton, B. W. Brunton, J. L. Proctor, E. Kaiser, and J. N. Kutz, "Chaos as an intermittently forced linear system," *arXiv preprint arXiv:1608.05306*, 2016.

Niall M. Mangan received the Dual B.S. degrees in mathematics and physics (with a minor in chemistry) from Clarkson University, Potsdam, NY, USA, in 2008, and the Ph.D. degree in systems biology from Harvard University, Cambridge, MA, USA, in 2013. She is currently an Acting Assistant Professor of Applied Mathematics with the University of Washington.

Steven L. Brunton (S'08–M'11) received the B.S. degree in mathematics (with a minor in control and dynamical systems) from the California Institute of Technology, Pasadena, CA, USA, in 2006, and the Ph.D. degree in mechanical and aerospace engineering from Princeton University, Princeton, NJ, USA, in 2012. He is currently an Assistant Professor of Mechanical Engineering and a Data Science Fellow with the eScience Institute, University of Washington.

Joshua L. Proctor received the B.S. degree in aeronautics and astronautics engineering and the B.A. degree in English literature from the University of Washington, Seattle, WA, USA, in 2006, and the Ph.D. degree in mechanical and aerospace engineering from Princeton University, Princeton, NJ, USA, in 2011. He is currently a Senior Research Scientist with the Institute for Disease Modeling and an Affiliate Assistant Professor of Applied Mathematics and Mechanical Engineering with the University of Washington.

J. Nathan Kutz received the B.S. degrees in physics and mathematics from the University of Washington, Seattle, WA, USA, in 1990, and the Ph.D. degree in applied mathematics from Northwestern University, Evanston, IL, USA, in 1994. He is currently a Professor of Applied Mathematics, an Adjunct Professor of Physics and Electrical Engineering, and a Senior Data Science Fellow with the eScience Institute, University of Washington.