

Volume 31, Issue 8, Aug. 2021

Finding nonlinear system equations and complex network structures from data: A sparse optimization approach

Chaos 31, 082101 (2021); doi.org/10.1063/5.0062042

Ying-Cheng Lai

Finding nonlinear system equations and complex network structures from data: A sparse optimization approach

Cite as: Chaos 31, 082101 (2021); doi: 10.1063/5.0062042
Submitted: 1 July 2021 · Accepted: 11 August 2021 ·
Published Online: 27 August 2021



Ying-Cheng Lai^{1,2,a)} 

AFFILIATIONS

¹School of Electrical, Computer, and Energy Engineering, Arizona State University, Tempe, Arizona 85287-5706, USA

²Department of Physics, Arizona State University, Tempe, Arizona 85287, USA

^{a)}Author to whom correspondence should be addressed: Ying-Cheng.Lai@asu.edu

ABSTRACT

In applications of nonlinear and complex dynamical systems, a common situation is that the system can be measured, but its structure and the detailed rules of dynamical evolution are unknown. The inverse problem is to determine the system equations and structure from time series. The principle of exploiting sparse optimization to find the equations of dynamical systems from data was first articulated in 2011 by the ASU group. The basic idea is to expand the system equations into a power series or a Fourier series of a finite number of terms and then to determine the vector of the expansion coefficients based solely on data through sparse optimization. This Tutorial presents a brief review of the recent progress in this area. Issues discussed include discovering the equations of stationary or nonstationary chaotic systems to enable the prediction of critical transition and system collapse, inferring the full topology of complex oscillator networks and social networks hosting evolutionary game dynamics, and identifying partial differential equations for spatiotemporal dynamical systems. Situations where sparse optimization works or fails are pointed out. The relation with the traditional delay-coordinate embedding method is discussed, and the recent development of a model-free, data-driven prediction framework based on machine learning is mentioned.

Published under an exclusive license by AIP Publishing. <https://doi.org/10.1063/5.0062042>

In experiments and real-world applications, the detailed equations of motion of an underlying system are often unknown, but observational data from the relevant dynamical variables of the system are available. As equations of motion are required for constructing a mathematical description of the dynamical system and thus represent a proof that some understanding of the system has been gained, a question is whether it would be possible to determine the system equations from data. This inverse problem of finding the system equations from data was first attempted by Crutchfield and McNamara in 1987, who developed an information based method to extract the effective equations of motion that may be different from the original equations. A method based on the inverse Frobenius–Perron operator to design a dynamical system that is “near” the original system in terms of the invariant density was proposed in 2000 by Bollt. About ten years ago, the idea of exploiting sparse optimization, e.g., compressive sensing, to find the governing equations of nonlinear dynamical systems was articulated by the ASU group (see Refs. 78 and 89). The method has been extended to finding the full connecting

topology of complex networks hosting oscillatory, evolutionary game, or binary-state dynamics and to finding partial differential equations (PDEs) for spatiotemporal dynamical systems. This Tutorial provides a succinct review of the sparse optimization principle for finding system equations from observational data. The range of applicability and limitations are discussed.

I. INTRODUCTION

In nonlinear dynamics, the traditional solution to the inverse problem, i.e., to analyze time series to probe into the inner “gears” of the system, is based on the paradigm of delay-coordinate embedding.^{1,2} The research started about four decades ago when Takens¹ proved that the underlying dynamical system can be faithfully reconstructed from a time series with a one-to-one correspondence between the reconstructed and the true but unknown dynamical systems. From the reconstructed system, statistical quantities characterizing the dynamical invariant set of the original system can

be assessed.^{3,4} For example, from a time series, the fractal dimensions of the underlying chaotic attractor can be estimated,^{5–12} as well as the Lyapunov exponents^{13–19} and some unstable periodic orbits.^{20–25} The continuity and differentiability of the original dynamical system can be tested.^{26–30} Practical issues on determining the basic parameters of delay-coordinate embedding such as the proper time delay^{11,12,31–36} and the embedding dimension³⁷ were addressed. Takens' paradigm was also extended to dynamical systems in the regime of transient chaos^{38–43} and to systems with a time delay.⁴⁴

There were previous methods on data-based identification and forecasting of nonlinear dynamical systems.^{45–68} One approach is to approximate a nonlinear system by a large collection of linear equations in different regions of the phase space to reconstruct the Jacobian matrices on a proper grid^{45,50,55} or to fit ordinary differential equations to chaotic data.⁵² Approaches based on chaotic synchronization⁵⁸ or genetic algorithms^{60,68} to system-parameter estimation were also investigated. In most studies, short-term predictions can be achieved. For nonstationary systems, the method of overembedding was introduced⁶² in which the time-varying parameters were treated as independent dynamical variables so that the essential aspects of determinism of the underlying system can be identified.

Takens' embedding paradigm, while having evolved into a powerful and effective framework in the past 40 years to address the inverse problem in nonlinear dynamical systems, gives only a topological equivalent of the system of interest: it does not give the equations of motion of the original system. As such, the state evolution cannot be predicted nor critical transitions leading to a possible system collapse upon parameter variations. The inverse problem of determining the system equations from measurements was addressed in 1987 by Crutchfield and McNamara,⁶⁹ who extended the notion of qualitative information contained in a sequence of observations to deduce the effective equations of motion that give the deterministic portion of the observed behavior. The effective equations, however, may contain additional terms that are not present in the original equations. The idea of using the inverse Frobenius–Perron problem through L^∞ to design a dynamical system that is “near” the original system with a desired invariant density was proposed in 2000 by Bollt.⁷⁰ Modeling and nonlinear parameter estimation using the least squares (L^2) best approximation (Kronecker product representation) were articulated and analyzed in 2007 by Yao and Bollt.⁷¹ In the past decade, the idea of using sparse (L^1) optimization methods such as compressive sensing^{72–77} was articulated^{78–86} for discovering the exact equations of motion for certain class of nonlinear and complex dynamical systems.⁸⁷ Quite recently, entropic regression for overcoming the problem of outliers in nonlinear system identification was articulated.⁸⁸

The principle of sparse optimization for finding the equations of nonlinear dynamical systems from data was first published⁷⁸ by the author's group at the Arizona State University in 2011.⁸⁹ The basic idea is that many nonlinear dynamical systems in nature and engineering are governed by smooth functions that can be approximated by series expansions. The inverse problem of determining the system equations then boils down to estimating the coefficients in the series. If the series contains many high-order terms, the total number of coefficients to be estimated will be large. In this case, there is no advantage to use the series expansions, and the

problem remains as difficult as with the original equations. However, if the system equations are relatively simple in the sense that most coefficients in the series expansions are zero, as in many classical nonlinear dynamical systems, the vector of all the coefficients to be determined will be sparse, rendering applicable and effective sparse optimization methods such as compressive sensing^{72,73,75–77} originally developed in the field of signal processing in engineering and applied mathematics. A virtue common to sparse optimization methods is the low requirement of observation data. In addition to enabling finding equations of nonlinear dynamical systems from data,^{78,90} compressive sensing has also been exploited for the reconstruction of complex networks with discrete and continuous time nodal dynamics^{78,80} and evolutionary game dynamics,⁷⁹ for detecting hidden nodes,^{82,84} for predicting and controlling network synchronization dynamics,⁸¹ and for reconstructing spreading dynamics based on binary data.⁸⁵

II. PRINCIPLE OF DISCOVERING SYSTEM EQUATIONS FROM DATA BASED ON SPARSE OPTIMIZATION

Compressive sensing solves the following convex optimization problem:

$$\min \|\mathbf{a}\|_1 \quad \text{subject to} \quad \mathcal{G} \cdot \mathbf{a} = \mathbf{X}, \quad (1)$$

where \mathbf{a} is a sparse vector to be determined, \mathcal{G} is a known random projection matrix, \mathbf{X} is a measurement vector constructed from the available data, and $\|\mathbf{a}\|_1 = \sum_{i=1}^N |\mathbf{a}_i|$ denotes the L_1 norm of vector \mathbf{a} . Compressive sensing is a paradigm of high-fidelity signal reconstruction using only sparse data,^{72,73,75–77} which was originally developed to solve the problem of transmitting massive data sets, such as those collected from a large-scale sensor network. Because of the high dimensionality, direct transmission of such data sets would require a broad bandwidth. However, a common situation with sensor networks is that, most of the time, majority of the sensors are inactive so that the data set collected from the entire network at a time is sparse. For example, say, a data set of N points is represented by an $N \times 1$ vector \mathbf{a} , where N is a large integer. Since \mathbf{a} is sparse, most of its entries are zero and only a small number of k entries are non-zero, where $k \ll N$. One can use a Gaussian random matrix \mathcal{G} of dimension $M \times N$ to obtain an $M \times 1$ vector \mathbf{X} : $\mathbf{X} = \mathcal{G} \cdot \mathbf{a}$, where $M \sim k$. Because the dimension of \mathbf{X} is much smaller than that of the original vector \mathbf{a} , transmitting \mathbf{X} would require a much smaller bandwidth, provided that \mathbf{a} can be reconstructed at the receiver end of the communication channel.

The problem of reconstructing the equations of a nonlinear dynamical system from data can be formulated in the compressive sensing framework.^{78,90} Consider a general dynamical system described by

$$\frac{d\mathbf{x}}{dt} = \mathbf{F}(\mathbf{x}), \quad (2)$$

where \mathbf{x} is an m -dimensional vector: $\mathbf{x} \equiv (x_1, x_2, \dots, x_m)^T$, and $\mathbf{F}(\mathbf{x})$ represents the velocity (vector) field of the system with m components: $\mathbf{F}(\mathbf{x}) = [F_1(\mathbf{x}), F_2(\mathbf{x}), \dots, F_m(\mathbf{x})]^T$. The goal is to determine $\mathbf{F}(\mathbf{x})$ from limited measured time series $\mathbf{x}(t)$. The basic idea⁷⁸ is to expand the velocity field as a multi-dimensional power series. In particular, the j th component of the velocity field can be expanded to

order q as

$$F_j(\mathbf{x}) = \sum_{l_1=0}^q \sum_{l_2=0}^q \dots \sum_{l_m=0}^q (a_j)_{l_1 l_2 \dots l_m} x_1^{l_1} x_2^{l_2} \dots x_m^{l_m}, \quad (3)$$

where $(a_j)_{l_1 l_2 \dots l_m}$ ($l_1, \dots, l_m = 1, \dots, q$) constitute the set of $(1 + q)^m$ coefficients to be determined from the measurements. If this set of coefficients is dense in the sense that most of them are nontrivial, then resorting to the power series expansion as in Eq. (3) does not lead to any step closer to the solution because the coefficients cannot be determined based on limited measurements. However, if the coefficient set is sparse in that most of its elements are zero, then it would be possible to use compressive sensing to uniquely solve for the nontrivial coefficients. Some well studied nonlinear dynamical systems, such as the classic Lorenz⁹¹ and Rössler⁹² oscillators, fall into this “sparse” category because their vector fields contain only a few power series terms.

To better understand the mathematical structure of the problem formulation, consider the concrete case of a three-dimensional phase space ($m = 3$) and a power series expansion up to order three ($n = 3$). In this case, the total number of unknown coefficients is $(1 + n)^m = 64$. For convenience, let the dynamical variables be $\mathbf{x}(t) \equiv [x(t), y(t), z(t)]^T$. The first component of the vector field can be written as

$$F_1(\mathbf{x}) = (a_1)_{000}x^0y^0z^0 + (a_1)_{100}x^1y^0z^0 + \dots + (a_1)_{333}x^3y^3z^3. \quad (4)$$

The $N = 64$ coefficients to be determined can be organized into a vector, a $N \times 1$ -dimensional vector,

$$\mathbf{a}_1 \equiv \begin{pmatrix} (a_1)_{000} \\ (a_1)_{100} \\ \dots \\ (a_1)_{333} \end{pmatrix}. \quad (5)$$

Defining all the combinations of the powers of the dynamical variables in Eq. (4) as a 1×64 -dimensional row vector,

$$\mathbf{g}(t) \equiv [x^0(t)y^0(t)z^0(t), x^1(t)y^0(t)z^0(t), \dots, x^3(t)y^3(t)z^3(t)], \quad (6)$$

we can write Eq. (4) as

$$F_1[\mathbf{x}(t)] = \mathbf{g}(t) \cdot \mathbf{a}_1. \quad (7)$$

Suppose measurements of the dynamical variables $\mathbf{x}(t)$ are available at $(M + 1)$ time instants t_0, t_1, \dots, t_M , where $M \ll N$. We have

$$\begin{aligned} dx(t_1)/dt &= F_1[\mathbf{x}(t_1)] = \mathbf{g}(t_1) \cdot \mathbf{a}_1 \\ dx(t_2)/dt &= F_1[\mathbf{x}(t_2)] = \mathbf{g}(t_2) \cdot \mathbf{a}_1 \\ &\dots \\ dx(t_M)/dt &= F_1[\mathbf{x}(t_M)] = \mathbf{g}(t_M) \cdot \mathbf{a}_1. \end{aligned} \quad (8)$$

The derivatives can be estimated from the measurements

$$\frac{dx(t_i)}{dt} \approx \frac{x(t_i) - x(t_{i-1})}{\delta t},$$

for $i = 1, \dots, M$. All the derivatives of M can be organized into a measurement vector that is $M \times 1$ -dimensional,

$$\mathbf{X} = \begin{pmatrix} dx(t_1)/dt \\ dx(t_2)/dt \\ \dots \\ dx(t_M)/dt \end{pmatrix}. \quad (9)$$

Likewise, the M row vectors $\mathbf{g}(t_1), \dots, \mathbf{g}(t_M)$, each being $1 \times N$ dimensional, can be organized into an $M \times N$ dimensional matrix

Measurement vector $\mathbf{X}_{M \times 1}$



=

Projection matrix $\mathcal{G}_{M \times N}$



$$k \leq M \ll N$$

Sparse coefficient vector $\mathbf{a}_{N \times 1}$ with k nonzero elements

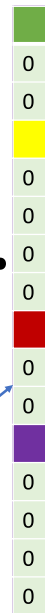


FIG. 1. Standard form of compressive sensing. The goal is to obtain the optimal solution of the N -dimensional coefficient vector \mathbf{a} from the M -dimensional measurement vector \mathbf{X} through the projection matrix \mathcal{G} that is $M \times N$ dimensional, where $M \ll N$. The mathematical framework of compressive sensing guarantees an optimal solution insofar as the coefficient vector to be solved is sparse: it has only k nontrivial elements, where $k \leq M$, provided that the projection matrix is random.

as

$$\mathcal{G} = \begin{pmatrix} \mathbf{g}(t_1) \\ \mathbf{g}(t_2) \\ \dots \\ \mathbf{g}(t_M) \end{pmatrix}. \tag{10}$$

Equation (8) can then be written as

$$\mathbf{X}_{M \times 1} = \mathcal{G}_{M \times N} \cdot (\mathbf{a}_1)_{N \times 1}. \tag{11}$$

The structure of Eq. (11) is shown in Fig. 1, where \mathcal{G} is the projection matrix. Suppose that the coefficient vector \mathbf{a}_1 is sparse: it has k non-trivial elements, where $k \ll N$. If the inequality $k \leq M \ll N$ holds, then Eq. (11) is in the standard form of compressive sensing,^{72,73,75-77} which is Eq. (1).

Two remarks are in order.

First, the coefficient vector \mathbf{a}_1 is associated with the velocity field of the first dynamical variable x . For any other dynamical variable in Eq. (2), a similar expansion procedure can be carried out. For example, for the second dynamical variable y , the coefficient vector \mathbf{a}_2 of its velocity field can be solved through

$$\mathbf{Y}_{M \times 1} = \mathcal{G}_{M \times N} \cdot (\mathbf{a}_2)_{N \times 1}, \tag{12}$$

where $\mathbf{Y}_{M \times 1}$ is the measurement vector constituting the derivatives of y at the measurement points. Note that the projection matrix $\mathcal{G}_{M \times N}$ has the same form in Eqs. (11) and (12).

Second, in the mathematical framework of compressive sensing, a requirement is that the projection matrix \mathcal{G} be random with zero correlation among its elements. However, in the power series expansion formulation, the elements of this matrix are distinct combinations of the powers of all the dynamical variables in the system. Since the system is deterministic, nonzero correlations among the matrix elements are inevitable. For chaotic systems, this violation of the randomness condition may not be as severe, since the time evolution of the dynamical variables is effectively random, insofar as the time interval between the adjacent measurement points is reasonably large. Nonetheless, there is no mathematical guarantee that an optimal solution can be obtained for solving the inverse problem in nonlinear dynamical systems through the compressive sensing approach, in spite of previously demonstrated successes.^{78-82,84,85,90}

III. APPLICATIONS

A number of applications of the compressive sensing based solutions of inverse problems in nonlinear and complex dynamical systems are described.

A. Predicting system collapse

When some parameter of a nonlinear dynamical system changes, a bifurcation that leads to the collapse of the system can occur. For example, in a global bifurcation called crisis,⁹³ at the bifurcation point, a chaotic attractor collides with its own basin boundary and is destroyed. Let p be the bifurcation parameter and p_c be the crisis point. Before the crisis, i.e., $p < p_c$, the system functions normally with a sustained chaotic behavior in its time evolution, as shown in Fig. 2(a), where the ordinate represents a typical dynamical variable of the system. Beyond the bifurcation point, the system collapses eventually after exhibiting transient chaos, as shown in Fig. 2(b). The

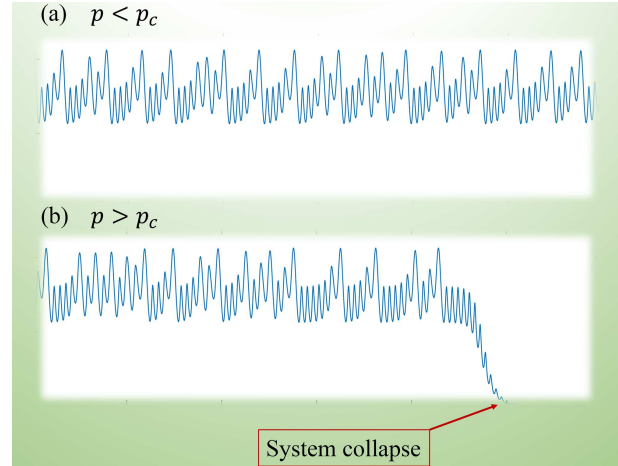


FIG. 2. Dynamical behaviors before and after a crisis bifurcation. The bifurcation occurs at the critical parameter value p_c . (a) Before the crisis ($p < p_c$), the system functions normally, as the average values of its dynamical variables maintain at a healthy level, in spite of chaotic fluctuations. (b) After the bifurcation, the system exhibits transient chaos and eventually collapses.

bifurcation is thus a catastrophe that must be prevented. In natural and engineering systems, catastrophic collapse is always a possibility. For example, in electrical power systems, voltage collapse⁹⁴ can occur after the system enters into the state of transient chaos.^{38,93} In ecology, slow parameter drift caused by environmental deterioration can induce a transition into transient chaos, followed by species extinction.^{95,96} For a dynamical system of interest, predicting a catastrophe in advance of its occurrence is of paramount importance. This is a challenging problem when the system equations are unknown and the only available information that one can rely on to make the prediction is time series measured while the system still functions normally.

If the underlying equations of the system have a simple mathematical structure, e.g., if its velocity field consists of a few power series or Fourier series terms only, then sparse optimization methods such as compressive sensing can be exploited to identify the system equations^{78,97} and consequently to predict transitions. In particular, based on the time series, one first predicts the system equations and then identifies the pertinent parameter of the system that can potentially lead to a collapse. With such information, one can perform a computational bifurcation analysis to locate potential catastrophic events in the parameter space so as to determine the likelihood of systems drifting into a catastrophic regime. For example, if it is determined that the system currently operates in a parameter region close to the crisis bifurcation, a catastrophe may be imminent as a small parameter change or a random perturbation can push the system into the regime of transient chaos where collapse is inevitable.

The principle of compressive sensing based prediction of system collapse has been demonstrated with a number of nonlinear dynamical systems.^{78,97}

B. Predicting future attractors in time-varying dynamical systems

Physical systems are constantly under external influences that can lead to parameter drifting. If the time scale of the internal dynamics of the system is much faster than that of the external perturbation, the resulting drift in the parameter can be regarded as adiabatic. In this case, in a time window of duration much longer than the internal but shorter than the external time scale, the system can be viewed to be in a type of “asymptotic state” or an “attractor,” as for the case of a stationary dynamical system with no time-varying parameters. However, in a long time scale, the attractor does depend on time. A problem of significant interest is to forecast the “future” attractor of the system. For example, consider the climate system. It is under random disturbances all the time, but adiabatic parameter drifting is also present, such as that induced by the injection of CO₂ into the atmosphere due to human activities. The time scale for any appreciable increase in the CO₂ level (e.g., months or years) is typically much slower than the intrinsic time scales of the system (e.g., days). The climate system can thus be regarded as an adiabatically time-varying, nonlinear dynamical system. To forecast the possible future attractor of the system is key to sustainability, an issue that is critical to many other natural and engineering systems.

It was demonstrated⁹⁰ that predicting the future attractor of an adiabatically time-varying dynamical system can be formulated as a problem solvable by compressive sensing. In particular, let the system be described by

$$dx/dt = F[x, p(t)], \tag{13}$$

where \mathbf{x} is the set of dynamical variables of the system in the m -dimensional phase space and $\mathbf{p}(t) \equiv [p_1(t), \dots, p_K(t)]$ denotes K independent, time-varying parameters. The tacit assumption is that both the velocity field \mathbf{F} and the vector parameter function $\mathbf{p}(t)$ are unknown, but only time series $\mathbf{x}(t)$ measured from the system in the time interval $t_M - T_M \leq t \leq t_M$ are available, where t_M is the current time. The goal is to determine the precise mathematical forms of \mathbf{F} and $\mathbf{p}(t)$ from the available time series at t_M so that the dynamical evolution of the system and the likely attractors for $t > t_M$ can be computationally assessed.

As for the case of predicting system collapse discussed in Sec. III A, the first step is to expand all components of the time-dependent vector field $F[\mathbf{x}, \mathbf{p}(t)]$ into a power series in terms of both dynamical variables \mathbf{x} and time t . The i th component $F[\mathbf{x}, \mathbf{p}(t)]_i$ of the vector field can be written as

$$\begin{aligned} & \sum_{l_1, \dots, l_m=1}^n [(\alpha_i)_{l_1, \dots, l_m} x_1^{l_1} \dots x_m^{l_m} \cdot \sum_{w=0}^v (\beta_i)_w t^w] \\ & \equiv \sum_{l_1, \dots, l_m=1}^n \sum_{w=0}^v (a_i)_{l_1, \dots, l_m; w} x_1^{l_1} \dots x_m^{l_m} \cdot t^w, \end{aligned} \tag{14}$$

where x_k ($k = 1, \dots, m$) is the k th component of the dynamical variable, a_i is the i th component of the coefficient vector to be determined, and the time evolution of each term can be approximated by the power series expansion in time, i.e., $\sum_{w=0}^v (\beta_i)_w t^w$. The power series expansion can then be cast into the standard form of

compressive sensing Eq. (1). If every combined scalar coefficient $(a_i)_{l_1, \dots, l_m; w}$ associated with the corresponding term in Eq. (14) can be determined from time series for $t \leq t_M$, the vector field component $[F(\mathbf{x}, \mathbf{p}(t))]_i$ becomes known. Repeating the procedure for all components, the entire vector field for $t > t_M$ can be identified.

Note that, the predicted form of \mathbf{F} and $\mathbf{p}(t)$ at time t_M would contain errors that, in general, will increase with time. In addition, for $t > t_M$, new perturbations can occur to the system so that the forms of \mathbf{F} and $\mathbf{p}(t)$ may be further changed. It is thus necessary to execute the prediction algorithm frequently using time series available at the time. For example, the system could be monitored at all times so that time series can be collected, and predictions can be carried out at t_i s, where $\dots > t_i > \dots > t_{M+2} > t_{M+1} > t_M$. For any t_i , the prediction algorithm is to be performed based on available time series in a suitable window prior to t_i .

C. Finding complex network structure from data

In complex dynamical networks, a typical inverse problem is to find the full connecting topology of the network based on observational time series from the nodes. If the network is sparse, as are many complex networks in the real world,⁹⁸ then compressive sensing can be used to address the inverse problem. In particular, given a sparsely connected network with unknown connecting structure and nodal dynamical equations, if oscillatory time series can be measured from all nodes in the network, it would be possible to determine the connection structure and identify the nodal equations of motion.⁸⁰

An oscillator network is a high-dimensional dynamical system that generates oscillatory time series at various nodes. The local dynamics at a node are described by

$$\dot{\mathbf{x}}_i = \mathbf{F}_i(\mathbf{x}_i) + \sum_{j=1, j \neq i}^n \mathcal{K}_{ij} \cdot (\mathbf{x}_j - \mathbf{x}_i), \quad (i = 1, \dots, n), \tag{15}$$

where $\mathbf{x}_i \in R^d$ represents the set of externally accessible dynamical variables of node i , n is the number of nodes, and \mathcal{K}_{ij} is the $d \times d$ coupling matrix between the dynamical variables at nodes i and j denoted by

$$\mathcal{K}_{ij} = \begin{pmatrix} k_{ij}^{1,1} & k_{ij}^{1,2} & \dots & k_{ij}^{1,d} \\ k_{ij}^{2,1} & k_{ij}^{2,2} & \dots & k_{ij}^{2,d} \\ \dots & \dots & \dots & \dots \\ k_{ij}^{d,1} & k_{ij}^{d,2} & \dots & k_{ij}^{d,d} \end{pmatrix}. \tag{16}$$

In the elements of matrix \mathcal{K}_{ij} , the superscripts lm ($l, m = 1, 2, \dots, d$) stand for the coupling from the k th component of the dynamical variable at node i to the l th component of the variable at node j . For any pair of nodes, the number of possible coupling terms is d^2 . If there is at least one nonzero element in matrix \mathcal{K}_{ij} , nodes i and j are coupled and, as a result, there is a link between them. Often, more than one element in \mathcal{K}_{ij} can be nonzero. However, if all the elements of \mathcal{K}_{ij} are zero, nodes i and j are not coupled. The connecting structure and the interaction strengths among various nodes of the network can be found if the coupling matrix \mathcal{K}_{ij} is determined from the observational data.

The first step in constructing a sparse optimization solution of the network inverse problem is to rewrite Eq. (15) as

$$\dot{\mathbf{x}}_i = [\mathbf{F}_i(\mathbf{x}_i) - \sum_{j=1, j \neq i}^n \mathcal{K}_{ij} \mathbf{x}_j] + \sum_{j=1, j \neq i}^n \mathcal{K}_{ij} \mathbf{x}_j, \quad (17)$$

where the first term in the square bracket is a function of the dynamical variable vector \mathbf{x}_i of node i and the second term is a function of variables of other nodes through node-to-node coupling. The terms in the square bracket can be conveniently denoted as $\mathbf{\Gamma}_i(\mathbf{x}_i)$. The l th component of $\mathbf{\Gamma}_i(\mathbf{x}_i)$ can be represented by a power series of order up to q

$$\begin{aligned} [\mathbf{\Gamma}_i(\mathbf{x}_i)]_l &\equiv \left[\mathbf{F}_i(\mathbf{x}_i) - \sum_{j=1, j \neq i}^n \mathcal{K}_{ij} \mathbf{x}_j \right]_l \\ &= \sum_{l_1=0}^q \sum_{l_2=0}^q \dots \sum_{l_d=0}^q [(\alpha_i)_{l_1, \dots, l_d}] (\mathbf{x}_i)_1^{l_1} (\mathbf{x}_i)_2^{l_2} \dots (\mathbf{x}_i)_d^{l_d}, \end{aligned} \quad (18)$$

where $(\mathbf{x}_i)_l$ ($l = 1, \dots, d$) is the l th component of the dynamical variable at node i , the total number of products is $(1 + q)^d$, and $[(\alpha_i)_{l_1, \dots, l_d}] \in R^m$ is the coefficient scalar of each product term, which is to be determined from measurements as well. Note that terms in Eq. (18) are all possible products of different components with different power of exponents. As an example, for $d = 2$ (the components are x and y) and $q = 2$, the power series expansion is

$$\begin{aligned} &\alpha_{0,0} + \alpha_{1,0}x + \alpha_{0,1}y + \alpha_{2,0}x^2 + \alpha_{0,2}y^2 + \alpha_{1,1}xy \\ &+ \alpha_{2,1}x^2y + \alpha_{1,2}xy^2 + \alpha_{2,2}x^2y^2. \end{aligned}$$

The second step is to rewrite Eq. (17) as

$$\dot{\mathbf{x}}_i = \mathbf{\Gamma}_i(\mathbf{x}_i) + \mathcal{K}_{i1} \mathbf{x}_1 + \mathcal{K}_{i2} \mathbf{x}_2 + \dots + \mathcal{K}_{in} \mathbf{x}_n. \quad (19)$$

The goal is to estimate the various coupling matrices \mathcal{K}_{ij} ($j = 1, \dots, i - 1, i + 1, \dots, n$) and the coefficients of $\mathbf{\Gamma}_i(\mathbf{x}_i)$ from sparse measurements. The sparsity requirement of compressive sensing stipulates that, to reconstruct the coefficients of Eq. (19) from data, most coefficients must be zero. To include as many coupling forms as possible, one can write each term $\mathcal{K}_{ij} \mathbf{x}_j$ in Eq. (19) as a power series in the same form of $\mathbf{\Gamma}_i(\mathbf{x}_i)$ but with different coefficients,

$$\dot{\mathbf{x}}_i = \mathbf{\Gamma}_1(\mathbf{x}_1) + \mathbf{\Gamma}_2(\mathbf{x}_2) + \dots + \mathbf{\Gamma}_n(\mathbf{x}_n). \quad (20)$$

This setting includes many possible coupling forms and ensures that the sparsity condition is met so that the prediction problem can be formulated in the compressive sensing framework. For an arbitrary node i , information about the node-to-node coupling, or about the network connectivity, is fully contained in $\mathbf{\Gamma}_j$ ($j \neq i$). For example, if in the equation of node i , a term in $\mathbf{\Gamma}_j$ ($j \neq i$) is not zero, nodes i and j are then coupled with the strength given by the coefficient of the term. Subtracting the coupling terms $-\sum_{j=1, j \neq i}^n \mathcal{K}_{ij} \mathbf{x}_j$ from $\mathbf{\Gamma}_i$ in Eq. (18), which is the sum of the coupling coefficients of all $\mathbf{\Gamma}_j$ ($j \neq i$), the local system equations $\mathbf{F}_i(\mathbf{x}_i)$ can be obtained. Once the coefficients of Eq. (20) have been determined, the nodal dynamical equations and the couplings among the nodes are all known. As explained in Sec. III A, the power series expansion coefficients of Eq. (20) can be determined through compressive sensing.

D. Finding social network structure from evolutionary game data

In the summer of 2011, a small social network experiment was conducted at the Arizona State University, where 22 students from different schools were invited to test the effectiveness of a sparse optimization based approach to mapping out the structure of social networks. In particular, the 22 participants constituted a social network, where any individual has a few acquaintances/friends in the group, but many participants had not known each other prior to the experiment. The network is thus sparse. During the experiment, all participants were asked to play an evolutionary game, the prisoner's dilemma game (PDG), with their friends for about 30 runs. That is, each and every agent (node) in the network was asked to play the game but only with his/her direct neighbors. For each run of the play, the strategies used by the opponents of each and every pair of players were recorded, together with the outcome (i.e., the winner and loser). Based on the data from all 30 runs, a compressive sensing based algorithm was executed, yielding a network structure that matches exactly with that of the actual social network. In fact, it was demonstrated that data from about 15 runs were already sufficient to infer the underlying network structure with 100% accuracy.⁷⁹

The theoretical underpinning of this successful experiment lies in formulating the dynamical process of evolutionary game on a network as a problem of sparse optimization. Specifically, in an evolutionary game, the players use different strategies in order to gain the maximum payoff, which, in general, can be divided into two types: cooperation and defection. It was shown⁷⁹ that, with limited data on each player's strategy and payoff, a compressive sensing based framework can be developed to yield precise knowledge of the node-to-node interaction patterns in an efficient manner.

A sketch of the principle underlying the compressive sensing framework is as follows. In an evolutionary game, at any time a player can choose one of the two strategies S : cooperation (C) or defection (D), which can be expressed as $\mathbf{S}(C) = (1, 0)^T$ and $\mathbf{S}(D) = (0, 1)^T$. The payoffs of two players in a game are determined by their strategies and the payoff matrix of the specific game. For example, for the prisoner's dilemma game (PDG)⁹⁹ and the snowdrift games (SG),¹⁰⁰ the payoff matrices are given, respectively, by

$$\mathcal{P}_{PDG} = \begin{pmatrix} 1 & 0 \\ b & 0 \end{pmatrix} \text{ and } \mathcal{P}_{SG} = \begin{pmatrix} 1 & 1-r \\ 1+r & 0 \end{pmatrix}, \quad (21)$$

where b ($1 < b < 2$) and r ($0 < r < 1$) are parameters characterizing the temptation to defect. When a defector encounters a cooperator, the defector gains payoff b in the PDG and payoff $1 + r$ in the SG, but the cooperator gains the sucker payoff 0 in the PDG and payoff $1 - r$ in the SG. At each time step, all agents play the game with their neighbors and gain payoffs. For agent i , the payoff is

$$P_i = \sum_{j \in \Gamma_i} \mathbf{S}_i^T \cdot \mathcal{P} \cdot \mathbf{S}_j, \quad (22)$$

where \mathbf{S}_i and \mathbf{S}_j are the strategies of agents i and j at the time and the sum is over the neighboring set Γ_i of i . After obtaining its payoff, an agent updates its strategy according to its own and neighbors' payoffs, attempting to maximize its payoff at the next round. Possible mathematical prescriptions to describe quantitatively an agent's decision making process include the best-take-over rule,⁹⁹ the Fermi

rule,¹⁰¹ and one based on the payoff-difference-determined updating probability.¹⁰² For example, the Fermi rule is defined, as follows. After player i randomly chooses a neighbor j , i adopts j 's strategy S_j with the probability¹⁰¹

$$W(S_i \leftarrow S_j) = \frac{1}{1 + \exp [(P_i - P_j)/\kappa]}, \quad (23)$$

where κ characterizes the stochastic uncertainties in the evolutionary game dynamics: $\kappa = 0$ corresponds to absolute rationality where the probability is 0 if $P_j < P_i$ and 1 if $P_i < P_j$, and $\kappa \rightarrow \infty$ indicates completely random decision making. The probability W thus characterizes the bounded rationality of agents in a society and the natural selection based on the relative fitness in evolution.

The key to solving the inverse problem of network reconstruction is the relationship between agents' payoffs and strategies. The interactions among the agents in the network can be characterized by an $n \times n$ adjacency matrix \mathcal{A} with elements $a_{ij} = 1$ if agents i and j are connected and $a_{ij} = 0$ otherwise. The payoff of agent x can be expressed by

$$P_x(t) = a_{x1} S_x^T(t) \cdot \mathcal{P} \cdot S_1(t) + \dots + a_{x,x-1} S_x^T(t) \cdot \mathcal{P} \cdot S_{x-1}(t) + a_{x,x+1} S_x^T(t) \cdot \mathcal{P} \cdot S_{x+1}(t) + \dots + a_{xn} S_x^T(t) \cdot \mathcal{P} \cdot S_n(t), \quad (24)$$

where a_{xi} ($i = 1, \dots, x-1, x+1, \dots, n$) represents a possible connection between agent x and its neighbor i , $a_{xi} S_x^T(t) \cdot \mathcal{P} \cdot S_i(t)$ ($i = 1, \dots, x-1, x+1, \dots, n$) stands for the possible payoff of agent x from playing the game with i (if there is no connection between x and i , the payoff is zero because $a_{xi} = 0$), and $t = 1, \dots, M$ is the number of rounds that all agents play the game with their neighbors. This relation provides a base to construct the vector \mathbf{X}_x and matrix \mathcal{G}_x in a proper compressive sensing framework to obtain the solution of the neighboring vector \mathbf{A}_x of agent x . In particular, one can define

$$\begin{aligned} \mathbf{X}_x &\equiv (P_x(t_1), P_x(t_2), \dots, P_x(t_M))^T, \\ \mathbf{A}_x &\equiv (a_{x1}, \dots, a_{x,x-1}, a_{x,x+1}, \dots, a_{xn})^T, \end{aligned} \quad (25)$$

and

$$\mathcal{G}_x \equiv \begin{pmatrix} F_{x1}(t_1) & \dots & F_{x,x-1}(t_1) & F_{x,x+1}(t_1) & \dots & F_{xn}(t_1) \\ F_{x1}(t_2) & \dots & F_{x,x-1}(t_2) & F_{x,x+1}(t_2) & \dots & F_{xn}(t_2) \\ \vdots & \dots & \vdots & \vdots & \vdots & \vdots \\ F_{x1}(t_M) & \dots & F_{x,x-1}(t_M) & F_{x,x+1}(t_M) & \dots & F_{xn}(t_M) \end{pmatrix},$$

where $F_{xy}(t_i) = S_x^T(t_i) \cdot \mathcal{P} \cdot S_y(t_i)$. The relation among the vectors \mathbf{X}_x , \mathbf{A}_x , and the matrix \mathcal{G}_x is given by exactly the same form as of compressive sensing Eq. (1),

$$\mathbf{X}_x = \mathcal{G}_x \cdot \mathbf{A}_x, \quad (26)$$

where \mathbf{A}_x is sparse due to the sparsity of the underlying network, making the compressive sensing framework applicable. Since $S_x^T(t_i)$ and $S_y(t_i)$ in $F_{xy}(t_i)$ come from data and \mathcal{P} is known, the vector \mathbf{X}_x can be obtained directly while the matrix \mathcal{G}_x can be calculated from the strategy and payoff data. The vector \mathbf{A}_x can thus be predicted based solely on the time series game data. Since the self-interaction terms a_{xx} are not included in the vector \mathbf{A}_x and the self-column

$[F_{xx}(t_1), \dots, F_{xx}(t_M)]^T$ is excluded from the matrix \mathcal{G}_x , the computation required for compressive sensing can be reduced. In a similar fashion, the neighboring vectors of all other agents can be predicted, yielding the network adjacency matrix

$$\mathcal{A} = (\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_n),$$

and, hence, the connection structure of the underlying social network.

E. Sparse optimization based on LASSO and applications in reconstructing complex networks with binary-state dynamics

A generalization of the compressive sensing approach to inverse problems in nonlinear and complex dynamical systems is sparse optimization based on LASSO (least absolute shrinkage and selection operator).^{103,104} In statistical and machine learning, LASSO is a regression method that embodies both variable selection and regularization to enhance the prediction accuracy and interpretability of the statistical model it produces.^{105,106} In particular, LASSO incorporates an L_1 -norm and an error control term to solve the sparse vector \mathbf{a} according to the constraint $\mathcal{G} \cdot \mathbf{a} = \mathbf{X}$ [as in Eq. (1)] from a small amount of data by optimizing

$$\min_{\mathbf{a}} \left\{ \frac{1}{2M} \|\mathcal{G} \cdot \mathbf{a} - \mathbf{X}\|_2^2 + \lambda \|\mathbf{a}\|_1 \right\}, \quad (27)$$

where $\|\mathbf{a}\|_1$ is the L_1 norm of \mathbf{X} assuring the sparsity of the solution, the least squares term $\|\mathcal{G} \cdot \mathbf{a} - \mathbf{X}\|_2^2$ guarantees the robustness of the solution against noise in the data, and λ is a nonnegative regularization parameter that affects the reconstruction performance in terms of the sparsity of the network, which can be determined by a cross-validation method.¹⁰⁷ The advantage of LASSO is similar to that of compressive sensing: the number M of bases (measurements) needed can be much less than the length of \mathbf{a} . The LASSO-based sparse optimization method was successfully applied to reconstructing the structures of complex networks.¹⁰³

In data-based reconstruction of complex networks, a difficult problem is when the nodal dynamical states are discrete and are of the binary type—a situation that arises commonly in nature, technology, and society.¹⁰⁸ In a networked system hosting binary nodal dynamics, each node can be in one of the two possible states, e.g., being active or inactive in neuronal and gene regulatory networks,¹⁰⁹ cooperation or defection in networks hosting evolutionary game dynamics,¹¹⁰ being susceptible or infected in epidemic spreading on social and technological networks,¹¹¹ two competing opinions in social communities,¹¹² etc. The interactions among the nodes are complex, and a state change can be triggered either deterministically (e.g., depending on the states of their neighbors) or randomly. Indeed, deterministic and stochastic state changes can account for a variety of emergent phenomena, such as the outbreak of epidemic spreading,¹¹³ cooperation among selfish individuals,¹¹⁴ oscillations in biological systems,¹¹⁵ power blackout,¹¹⁶ financial crisis,¹¹⁷ and phase transitions in natural systems.¹¹⁸ A variety of models have been introduced to gain insights into binary-state dynamics on complex networks,⁹⁸ such as the voter models for competition of two opinions,¹¹⁹ stochastic propagation models for epidemic spreading,¹²⁰ models of rumor diffusion and adoption of

new technologies,¹²¹ cascading failure models,¹²² Ising spin models for ferromagnetic phase transition,¹²³ and evolutionary games for cooperation and altruism.¹⁰² A general theoretical approach to dealing with networks hosting binary state dynamics was developed¹²⁴ based on the pair approximation and the master equations, providing a good understanding of the effect of the network structure on the emergent phenomena.

The problem of reconstructing complex networks with binary-state dynamics is challenging for three reasons. First, the basic nodal dynamics are governed by the probability to transition between the two states—the switching probability of a node, which depends on the states of its neighbors according to a variety of functions that can be linear, nonlinear, piecewise, or stochastic. If the function that governs the switching probability is unknown, it would be difficult to obtain a solution of the reconstruction problem. Second, structural information is often hidden in the binary states of the nodes in an unknown manner and the dimension of the solution space can be high, rendering impractical (computationally prohibitive) brute-force enumeration of all possible network configurations. Third, the presence of measurement noise, missing data, and stochastic effects in the switching probability make the reconstruction task even more challenging, calling for the development of effective methods that are robust against internal and external random effects.

In 2017, a general and robust framework for reconstructing complex networks based solely on the binary states of the nodes without any knowledge about the switching functions was developed.¹⁰⁴ The idea was centered about developing a general method to linearize the switching functions from binary data. The data-based linearization method was demonstrated to be applicable to linear, nonlinear, piecewise, or stochastic switching functions. The method allows one to convert the network reconstruction problem into a sparse signal reconstruction problem for local structures associated with each node. In particular, because of the natural sparsity of complex networks, LASSO was used^{103,104} to identify the neighbors of each node in the network from sparse binary data contaminated by noise. The linearization procedure was justified through a number of linear, nonlinear, and piecewise binary-state dynamics on a large number of model and real complex networks. For all the models tested, universally high reconstruction accuracy was achieved¹⁰⁴ even for small data amount with noise. Because of its high accuracy, efficiency, and robustness against noise and missing data, the reconstruction framework can serve as a general solution to the inverse problem of network reconstruction from binary-state time series, which is key to articulating effective strategies to control complex networks with binary state dynamics.

F. Discovering models of spatiotemporal dynamical systems from data

There was an early work on modeling and parameter estimation for coupled oscillators and spatiotemporal dynamical systems based on the Kronecker product presentation.⁷¹ In recent years, sparse optimization or learning has been applied to discovering models of spatiotemporal systems described by partial differential equations (PDEs).^{125–129} PDEs for spatiotemporal systems in science

and engineering have the general form

$$\sum_{i=0}^n c_i \mathbf{f}_i(\mathbf{u}, \partial_t \mathbf{u}, \nabla \mathbf{u}, \nabla^2 \mathbf{u}, \dots) = 0, \tag{28}$$

where c_i s are constant coefficients and \mathbf{f}_i s are vector functions of time and space derivatives of various orders of the vector field \mathbf{u} . Usually, symmetry and physical considerations can be used to reduce the number of terms in Eq. (28) and to narrow down the possible functional forms of \mathbf{f}_i . From the measurement of (noisy) spatiotemporal data \mathbf{u} , sparse optimization can be used to remove the unnecessary terms and yield a model that contains a small number of terms.^{125–129}

In Ref. 129, the following procedure was devised. Consider a system that contains only a single term of the first-order time derivative of the vector field, written as

$$\partial_t \mathbf{u} = \sum_{i=1}^n c_i \mathbf{f}_i(\mathbf{u}, \nabla \mathbf{u}, \nabla^2 \mathbf{u}, \dots). \tag{29}$$

To obtain a set of linear equations for sparse optimization, one multiplies a weight vector \mathbf{w} with Eq. (29) and integrates both sides over a number of distinct spatiotemporal domains Ω_l ($l = 1, \dots, L$) to convert Eq. (29) to

$$\mathbf{q}_0 = \sum_{i=1}^n c_i \mathbf{q}_i = \mathcal{Q} \cdot \mathbf{c}, \tag{30}$$

where $\mathbf{c} \equiv (c_1, \dots, c_n)^T$ is the coefficient vector to be determined from data, \mathbf{q}_i is an L -dimensional column vector with entries given by

$$q_i^l = \int_{\Omega_l} \mathbf{q} \cdot \mathbf{f}_i d\Omega, \tag{31}$$

and $\mathcal{Q} \equiv (\mathbf{q}_1, \dots, \mathbf{q}_n)$ constitutes the “library” of possible terms \mathbf{q}_i s. Note that the integral in Eq. (31) involves derivatives of the vector field \mathbf{u} . If the measurements of \mathbf{u} are noisy, evaluating these derivatives can be problematic. However, performing integration by parts entails transferring the derivatives to the weight vector \mathbf{w} , which can be chosen to be smooth.

For $L \geq n$, an iterative sparse regression algorithm can be used to solve¹²⁹ the coefficient vector in Eq. (31). Each iteration involves the following form of the solution that minimizes the residual in Eq. (31),

$$\bar{\mathbf{c}} = \mathcal{Q}^+ \cdot \mathbf{q}_0, \tag{32}$$

where \mathcal{Q}^+ is the pseudoinverse of the matrix \mathcal{Q} . Using an empirical thresholding procedure to eliminate dynamically irrelevant terms \mathbf{q}_i , with the solved sparse coefficient vector \mathbf{c} , one can obtain a minimal PDE model from the measurements \mathbf{u} . The method was tested¹²⁹ with the one-dimensional Kuramoto–Sivashinsky model that has in its solutions spatiotemporal chaos, the two-dimensional Navier–Stokes equation, and the reaction–diffusion equation.

IV. DISCUSSION

The principle of exploiting sparse optimization such as compressive sensing to find the equations of nonlinear dynamical systems from data was first articulated⁷⁸ in 2011. The basic idea is

to expand the equations (the velocity field for a continuous time dynamical system or the map function for a discrete time system) of the underlying system into a power series or a Fourier series of a finite number of terms and then to determine the vector of the expansion coefficients based on data through sparse optimization. The sparse optimization principle has been demonstrated to be effective for finding the governing equations of certain types of nonlinear dynamical systems for inferring the detailed connection structures of complex dynamical networks such as oscillator networks and social networks hosting evolutionary game dynamics. In spite of the demonstrated success, limitations and open questions remain.

A key requirement is that the coefficient vector to be determined must be sparse. If the vector field or the map function contains a few power series terms, such as the classical Lorenz⁹¹ or Rössler⁹² chaotic oscillator, or contains a few Fourier series terms, such as the standard map,^{130,131} then sparse optimization can be quite effective and computationally efficient for finding the system equations.⁷⁸ However, if the vector field or the map function contains a large number of terms in its power series or Fourier series expansion so that the coefficient vector to be determined is dense, then the sparse optimization methodology will fail. One such example is the classical Ikeda map^{132,133} that describes the propagation of a laser pulse in an optical cavity,

$$\mathbf{F}(x, y) = \begin{pmatrix} a + b(x \cos \phi - y \sin \phi) \\ b(x \sin \phi + y \cos \phi) \end{pmatrix}, \quad (33)$$

with the nonlinear phase variable ϕ given by

$$\phi \equiv p - \frac{k}{1 + x^2 + y^2}, \quad (34)$$

where a , b , k , and p are parameters. It can be seen that both components of the map function contain an infinite number of power series terms, rendering inapplicable sparse optimization for finding the system equations from data.

In the mathematical formulation of compressive sensing Eq. (1), a requirement is that the projection matrix \mathcal{G} be random, e.g., Gaussian type of random matrices with no correlations among the matrix elements.^{72,73,75–77} However, in the power series formulation, e.g., Eq. (11), the elements of the projection matrix are different combinations of the powers of the dynamical variables, which are correlated even for a chaotic system. The demonstrated success in finding the system equations as reviewed in this Tutorial thus has no mathematical guarantee. It may also be possible that the “workable” domain of sparse optimization is larger than that guaranteed by rigorous mathematics. To possibly relax the conditions under which sparse optimization is effective remains an open mathematical issue.

Another difficulty with the application of sparse optimization to find system equations from data is the need to collect time series from all dynamical variables of the system. In real-world applications, situations are common where only a limited set of the intrinsic dynamical variables of the system are externally accessible. The requirement of observing all dynamical variables thus represents a formidable obstacle to the actual application of the

sparse optimization methods. This should be contrasted to the traditional delay-coordinate embedding paradigm^{1,2} where, in principle, measurements from a single dynamical variable are sufficient to reconstruct the phase space of the underlying system. The capability of the embedding paradigm in uncovering the topological and statistical properties of the dynamical invariant set responsible for the observed data notwithstanding, it is unable to yield the system equations.

In the past several years, machine learning has emerged as a promising paradigm for predicting the state evolution of nonlinear dynamical systems. In particular, a class of recurrent neural networks,^{134–137} the so-called reservoir computing machines, have attracted considerable attention since 2017 as a powerful paradigm for model-free, fully data driven prediction of nonlinear and chaotic dynamical systems.^{138–155} A typical reservoir computing machine consists of an input layer, a hidden layer that is usually a complex dynamical network, and an output layer. Time series data from the dynamical system to be predicted are used to train the machine through a series of adjustments to the weights that connect the hidden layer with the output layer. Once the machine has been trained, it can predict the state evolution of the target system for certain duration of time. A well trained reservoir computing machine can thus be viewed as a “replica” of the target system, where temporal synchronization between the two can be maintained.¹⁵² While the machine learning approach does not yield the equations of the system, it can be used to predict the system behavior especially for those that do not meet the sparsity condition. For example, it has been demonstrated recently¹⁵⁵ that crisis and transient chaos in the Ikeda map system (33) can be predicted by parameter-cognizant reservoir computing machines.¹⁵⁴

ACKNOWLEDGMENTS

This work was supported by the ONR under Grant No. N00014-21-1-2323. I thank Prof. Wenxu Wang (a postdoctoral fellow at ASU from 2008 to 2011) for initiating the research on the sparse-optimization approach to finding system equations and inferring complex network topology from data. I would also like to thank my former PhD students involved in this research: Dr. Rui Yang, Dr. Riqi Su, and Dr. Xuan Ni.

DATA AVAILABILITY

The data that support the findings of this study are available within the article.

REFERENCES

- ¹F. Takens, “Detecting strange attractors in fluid turbulence,” in *Dynamical Systems and Turbulence*, Lecture Notes in Mathematics Vol. 898, edited by D. Rand and L. S. Young (Springer-Verlag, Berlin, 1981), pp. 366–381.
- ²N. H. Packard, J. P. Crutchfield, J. D. Farmer, and R. S. Shaw, “Geometry from a time series,” *Phys. Rev. Lett.* **45**, 712–716 (1980).
- ³H. Kantz and T. Schreiber, *Nonlinear Time Series Analysis*, 1st ed. (Cambridge University Press, Cambridge, 1997).
- ⁴R. Hegger, H. Kantz, and R. Schreiber, *TISEAN*, e-book ed., (Max Planck Institute for Physics of Complex Systems, Dresden, 2007), see http://www.mpi-pks-dresden.mpg.de/tisean/TISEAN_3.01/index.html.
- ⁵P. Grassberger and I. Procaccia, “Measuring the strangeness of strange attractors,” *Physica D* **9**, 189–208 (1983).

- ⁶P. Grassberger, "Do climatic attractors exist?," *Nature* **323**, 609–612 (1986).
- ⁷I. Procaccia, "Complex or just complicated?," *Nature* **333**, 498–499 (1988).
- ⁸A. Osborne and A. Provenzale, "Finite correlation dimension for stochastic systems with power-law spectra," *Physica D* **35**, 357–381 (1989).
- ⁹E. Lorenz, "Dimension of weather and climate attractors," *Nature* **353**, 241–244 (1991).
- ¹⁰M. Ding, C. Grebogi, E. Ott, T. D. Sauer, and J. A. Yorke, "Plateau onset for correlation dimension: When does it occur?," *Phys. Rev. Lett.* **70**, 3872–3875 (1993).
- ¹¹Y.-C. Lai, D. Lerner, and R. Hayden, "An upper bound for the proper delay time in chaotic time series analysis," *Phys. Lett. A* **218**, 30–34 (1996).
- ¹²Y.-C. Lai and D. Lerner, "Effective scaling regime for computing the correlation dimension in chaotic time series analysis," *Physica D* **115**, 1–18 (1998).
- ¹³A. Wolf, J. B. Swift, H. L. Swinney, and J. A. Vastano, "Determining Lyapunov exponents from a time series," *Physica D* **16**, 285–317 (1985).
- ¹⁴M. Sano and Y. Sawada, "Measurement of the Lyapunov spectrum from a chaotic time series," *Phys. Rev. Lett.* **55**, 1082–1085 (1985).
- ¹⁵J. P. Eckmann and D. Ruelle, "Ergodic theory of chaos and strange attractors," *Rev. Mod. Phys.* **57**, 617–656 (1985).
- ¹⁶J. P. Eckmann, S. O. Kamphorst, D. Ruelle, and S. Ciliberto, "Lyapunov exponents from time series," *Phys. Rev. A* **34**, 4971–4979 (1986).
- ¹⁷R. Brown, P. Bryant, and H. D. I. Abarbanel, "Computing the Lyapunov spectrum of a dynamical system from an observed time series," *Phys. Rev. A* **43**, 2787–2806 (1991).
- ¹⁸T. D. Sauer, J. A. Tempkin, and J. A. Yorke, "Spurious Lyapunov exponents in attractor reconstruction," *Phys. Rev. Lett.* **81**, 4341–4344 (1998).
- ¹⁹T. D. Sauer and J. A. Yorke, "Reconstructing the Jacobian from data with observational noise," *Phys. Rev. Lett.* **83**, 1331–1334 (1999).
- ²⁰D. P. Lathrop and E. J. Kostelich, "Characterization of an experimental strange attractor by periodic orbits," *Phys. Rev. A* **40**, 4028–4031 (1989).
- ²¹R. Badii, E. Brun, M. Finardi, L. Flepp, R. Holzner, J. Pariso, C. Reyl, and J. Simonet, "Progress in the analysis of experimental chaos through periodic orbits," *Rev. Mod. Phys.* **66**, 1389–1415 (1994).
- ²²D. Pierson and F. Moss, "Detecting periodic unstable points in noisy chaotic and limit-cycle attractors with applications to biology," *Phys. Rev. Lett.* **75**, 2124–2127 (1995).
- ²³X. Pei and F. Moss, "Characterization of low-dimensional dynamics in the crayfish caudal photoreceptor," *Nature* **379**, 618–621 (1996).
- ²⁴P. So, E. Ott, S. J. Schiff, D. T. Kaplan, T. D. Sauer, and C. Grebogi, "Detecting unstable periodic orbits in chaotic experimental data," *Phys. Rev. Lett.* **76**, 4705–4708 (1996).
- ²⁵S. Allie and A. Mees, "Finding periodic points from short time series," *Phys. Rev. E* **56**, 346–350 (1997).
- ²⁶L. M. Pecora, T. L. Carroll, and J. F. Heagy, "Statistics for mathematical properties of maps between time series embeddings," *Phys. Rev. E* **52**, 3420–3439 (1995).
- ²⁷L. M. Pecora and T. L. Carroll, "Discontinuous and nondifferentiable functions and dimension increase induced by filtering chaotic data," *Chaos* **6**, 432–439 (1996).
- ²⁸L. M. Pecora, T. L. Carroll, and J. F. Heagy, "Statistics for continuity and differentiability: An application to attractor reconstruction from time series," *Fields Inst. Commun.* **11**, 49–62 (1997).
- ²⁹C. L. Goodridge, L. M. Pecora, T. L. Carroll, and F. J. Rachford, "Detecting functional relationships between simultaneous time series," *Phys. Rev. E* **64**, 026221 (2001).
- ³⁰L. M. Pecora, L. Moniz, J. Nichols, and T. Carroll, "A unified approach to attractor reconstruction," *Chaos* **17**, 013110 (2007).
- ³¹J. Theiler, "Spurious dimension from correlation algorithms applied to limited time series data," *Phys. Rev. A* **34**, 2427–2432 (1986).
- ³²W. Liebert and H. G. Schuster, "Proper choice of the time-delay for the analysis of chaotic time-series," *Phys. Lett. A* **142**, 107–111 (1989).
- ³³W. Liebert, K. Pawelzik, and H. G. Schuster, "Optimal embeddings of chaotic attractors from topological considerations," *Europhys. Lett.* **14**, 521–526 (1991).
- ³⁴T. Buzug and G. Pfister, "Optimal delay time and embedding dimension for delay-time coordinates by analysis of the global static and local dynamic behavior of strange attractors," *Phys. Rev. A* **45**, 7073–7084 (1992).
- ³⁵G. Kember and A. C. Fowler, "A correlation-function for choosing time delays in-phase portrait reconstructions," *Phys. Lett. A* **179**, 72–80 (1993).
- ³⁶M. T. Rosenstein, J. J. Collins, and C. J. D. Luca, "Reconstruction expansion as a geometry-based framework for choosing proper delay times," *Physica D* **73**, 82–98 (1994).
- ³⁷T. D. Sauer, J. A. Yorke, and M. Casdagli, "Embedology," *J. Stat. Phys.* **65**, 579–616 (1991).
- ³⁸Y.-C. Lai and T. Tél, *Transient Chaos—Complex Dynamics on Finite Time Scales*, 1st ed. (Springer, New York, 2011).
- ³⁹J. János, L. Flepp, and T. Tél, "Exploring transient chaos in an NMR-laser experiment," *Phys. Rev. Lett.* **73**, 529–532 (1994).
- ⁴⁰J. János and T. Tél, "Time-series analysis of transient chaos," *Phys. Rev. E* **49**, 2756–2763 (1994).
- ⁴¹M. Dhamala, Y.-C. Lai, and E. J. Kostelich, "Detecting unstable periodic orbits from transient chaotic time series," *Phys. Rev. E* **61**, 6485–6489 (2000).
- ⁴²M. Dhamala, Y.-C. Lai, and E. J. Kostelich, "Analysis of transient chaotic time series," *Phys. Rev. E* **64**, 056207 (2001).
- ⁴³I. Triandaf, E. Bollt, and I. B. Schwartz, "Approximating stable and unstable manifolds in experiments," *Phys. Rev. E* **67**, 037201 (2003).
- ⁴⁴S. R. Taylor and S. A. Campbell, "Approximating chaotic saddles in delay differential equations," *Phys. Rev. E* **75**, 046215 (2007).
- ⁴⁵J. D. Farmer and J. J. Sidorowich, "Predicting chaotic time series," *Phys. Rev. Lett.* **59**, 845–848 (1987).
- ⁴⁶M. Casdagli, "Nonlinear prediction of chaotic time series," *Physica D* **35**, 335–356 (1989).
- ⁴⁷G. Sugihara, B. T. Grenfell, and R. M. May, "Distinguishing error from chaos in ecological time series," *Philos. Trans. R. Soc. B* **330**, 235–251 (1990).
- ⁴⁸J. Kurths and A. A. Ruzmaikin, "On forecasting the sunspot numbers," *Solar Phys.* **126**, 407–410 (1990).
- ⁴⁹P. Grassberger and T. Schreiber, "Nonlinear time sequence analysis," *Int. J. Bifurcation Chaos* **1**, 521–547 (1990).
- ⁵⁰G. Gouesbet, "Reconstruction of standard and inverse vector fields equivalent to a Rössler system," *Phys. Rev. A* **44**, 6264–6280 (1991).
- ⁵¹A. A. Tsonis and J. B. Elsner, "Nonlinear prediction as a way of distinguishing chaos from random fractal sequences," *Nature* **358**, 217–220 (1992).
- ⁵²E. Baake, M. Baake, H. G. Bock, and K. M. Briggs, "Fitting ordinary differential equations to chaotic data," *Phys. Rev. A* **45**, 5524–5529 (1992).
- ⁵³A. Longtin, "Nonlinear forecasting of spike trains from sensory neurons," *Int. J. Bifurcation Chaos* **3**, 651–661 (1993).
- ⁵⁴D. B. Murray, "Forecasting a chaotic time series using an improved metric for embedding space," *Physica D* **68**, 318–325 (1993).
- ⁵⁵T. Sauer, "Reconstruction of dynamical systems from interspike intervals," *Phys. Rev. Lett.* **72**, 3811–3814 (1994).
- ⁵⁶G. Sugihara, "Nonlinear forecasting for the classification of natural time series," *Philos. Trans. R. Soc. A* **348**, 477–495 (1994).
- ⁵⁷B. Finkenstädt and P. Kuhbier, "Forecasting nonlinear economic time series: A simple test to accompany the nearest neighbor approach," *Empirical Econ.* **20**, 243–263 (1995).
- ⁵⁸U. Parlitz, "Estimating model parameters from time series by autosynchronization," *Phys. Rev. Lett.* **76**, 1232 (1996).
- ⁵⁹S. J. Schiff, P. So, T. Chang, R. E. Burke, and T. Sauer, "Detecting dynamical interdependence and generalized synchrony through mutual prediction in a neural ensemble," *Phys. Rev. E* **54**, 6708–6724 (1996).
- ⁶⁰G. G. Szpiro, "Forecasting chaotic time series with genetic algorithms," *Phys. Rev. E* **55**, 2557–2568 (1997).
- ⁶¹R. Hegger, H. Kantz, and T. Schreiber, "Practical implementation of nonlinear time series methods: The TISEAN package," *Chaos* **9**, 413–435 (1999).
- ⁶²R. Hegger, H. Kantz, L. Matassini, and T. Schreiber, "Coping with nonstationarity by overembedding," *Phys. Rev. Lett.* **84**, 4092–4095 (2000).
- ⁶³S. Sello, "Solar cycle forecasting: A nonlinear dynamics approach," *Astron. Astrophys.* **377**, 312–320 (2001).
- ⁶⁴T. Matsumoto, Y. Nakajima, M. Saito, J. Sugi, and H. Hamagishi, "Reconstructions and predictions of nonlinear dynamical systems: A hierarchical Bayesian approach," *IEEE Trans. Signal Proc.* **49**, 2138–2155 (2001).
- ⁶⁵L. A. Smith, "What might we learn from climate forecasts?," *Proc. Natl. Acad. Sci. U.S.A.* **19**, 2487–2492 (2002).

- ⁶⁶K. Judd, "Nonlinear state estimation, indistinguishable states, and the extended kalman filter," *Physica D* **183**, 273–281 (2003).
- ⁶⁷T. D. Sauer, "Reconstruction of shared nonlinear dynamics in a network," *Phys. Rev. Lett.* **93**, 198701 (2004).
- ⁶⁸C. Tao, Y. Zhang, and J. J. Jiang, "Estimating system parameters from chaotic time series with synchronization optimized by a genetic algorithm," *Phys. Rev. E* **76**, 016209 (2007).
- ⁶⁹J. P. Crutchfield and B. McNamara, "Equations of motion from a data series," *Complex Syst.* **1**, 417–452 (1987).
- ⁷⁰E. M. Bollt, "Controlling chaos and the inverse Frobenius-Perron problem: Global stabilization of arbitrary invariant measures," *Int. J. Bifurcation Chaos* **10**, 1033–1050 (2000).
- ⁷¹C. Yao and E. M. Bollt, "Modeling and nonlinear parameter estimation with Kronecker product representation for coupled oscillators and spatiotemporal systems," *Physica D* **227**, 78–99 (2007).
- ⁷²E. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory* **52**, 489–509 (2006).
- ⁷³E. Candès, J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Commun. Pure Appl. Math.* **59**, 1207–1223 (2006).
- ⁷⁴E. Candès, "Compressive sampling," in *Proceedings of the International Congress of Mathematicians* (Madrid, Spain, 2006), Vol. 3, pp. 1433–1452.
- ⁷⁵D. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory* **52**, 1289–1306 (2006).
- ⁷⁶R. G. Baraniuk, "Compressed sensing," *IEEE Signal Process. Mag.* **24**, 118–121 (2007).
- ⁷⁷E. Candès and M. Wakin, "An introduction to compressive sampling," *IEEE Signal Process. Mag.* **25**, 21–30 (2008).
- ⁷⁸W.-X. Wang, R. Yang, Y.-C. Lai, V. Kovanis, and C. Grebogi, "Predicting catastrophes in nonlinear dynamical systems by compressive sensing," *Phys. Rev. Lett.* **106**, 154101 (2011).
- ⁷⁹W.-X. Wang, Y.-C. Lai, C. Grebogi, and J.-P. Ye, "Network reconstruction based on evolutionary-game data via compressive sensing," *Phys. Rev. X* **1**, 021021 (2011).
- ⁸⁰W.-X. Wang, R. Yang, Y.-C. Lai, V. Kovanis, and M. A. F. Harrison, "Time-series-based prediction of complex oscillator networks via compressive sensing," *Europhys. Lett.* **94**, 48006 (2011).
- ⁸¹R.-Q. Su, X. Ni, W.-X. Wang, and Y.-C. Lai, "Forecasting synchronizability of complex networks from data," *Phys. Rev. E* **85**, 056220 (2012).
- ⁸²R.-Q. Su, W.-X. Wang, and Y.-C. Lai, "Detecting hidden nodes in complex networks from time series," *Phys. Rev. E* **85**, 065201 (2012).
- ⁸³R.-Q. Su, Y.-C. Lai, and X. Wang, "Identifying chaotic FitzHugh-Nagumo neurons using compressive sensing," *Entropy* **16**, 3889–3902 (2014).
- ⁸⁴R.-Q. Su, Y.-C. Lai, X. Wang, and Y.-H. Do, "Uncovering hidden nodes in complex networks in the presence of noise," *Sci. Rep.* **4**, 3944 (2014).
- ⁸⁵Z. Shen, W.-X. Wang, Y. Fan, Z. Di, and Y.-C. Lai, "Reconstructing propagation networks with natural diversity and identifying hidden sources," *Nat. Commun.* **5**, 4323 (2014).
- ⁸⁶R.-Q. Su, W.-W. Wang, X. Wang, and Y.-C. Lai, "Data based reconstruction of complex geospatial networks, nodal positioning, and detection of hidden node," *R. Soc. Open Sci.* **3**, 150577 (2016).
- ⁸⁷Professor E. M. Bollt from Clarkson University conceived the same idea of exploiting sparse optimization for discovering system equations from data, private communication (2019).
- ⁸⁸A. A. R. AlMomani, S. Jie, and E. M. Bollt, "How entropic regression beats the outliers problem in nonlinear system identification," *Chaos* **30**, 013107 (2020).
- ⁸⁹The idea of using sparse optimization to discover system equations was also published in 2016 [S. L. Brunton, J. L. Proctor, and J. Nathan Kutz, "Discovering governing equations from data by sparse identification of nonlinear dynamical systems," *Proc. Natl. Acad. Sci. U.S.A.* **113**, 3932–3937 (2016)] by Professor Kutz's group at the University of Washington. On October 8, 2015, Professor Kutz gave a talk at an AFOSR (Air Force Office of Scientific Research) Program Review meeting about this idea. The author of the present manuscript (Y.-C.L.) was in the audience and rose to point out politely that the idea had already been published in 2011 by the ASU group, and he immediately forwarded the 2011 paper to Professor Kutz.
- ⁹⁰R. Yang, Y.-C. Lai, and C. Grebogi, "Forecasting the future: Is it possible for time-varying nonlinear dynamical systems?," *Chaos* **22**, 033119 (2012).
- ⁹¹E. N. Lorenz, "Deterministic nonperiodic flow," *J. Atmos. Sci.* **20**, 130–141 (1963).
- ⁹²O. E. Rössler, "Equation for continuous chaos," *Phys. Lett. A* **57**, 397–398 (1976).
- ⁹³C. Grebogi, E. Ott, and J. A. Yorke, "Crises, sudden changes in chaotic attractors and chaotic transients," *Physica D* **7**, 181–200 (1983).
- ⁹⁴M. Dhamala and Y.-C. Lai, "Controlling transient chaos in deterministic flows with applications to electrical power systems and ecology," *Phys. Rev. E* **59**, 1646–1655 (1999).
- ⁹⁵K. McCann and P. Yodzis, "Nonlinear dynamics and population disappearances," *Am. Nat.* **144**, 873–879 (1994).
- ⁹⁶A. Hastings, K. C. Abbott, K. Cuddington, T. Francis, G. Gellner, Y.-C. Lai, A. Morozov, S. Petrivskii, K. Scranton, and M. L. Zeeman, "Transient phenomena in ecology," *Science* **361**, eaat6412 (2018).
- ⁹⁷W. Wang, Y.-C. Lai, and C. Grebogi, "Data based identification and prediction of nonlinear and complex dynamical systems," *Phys. Rep.* **644**, 1–76 (2016).
- ⁹⁸M. E. J. Newman, *Networks: An Introduction* (Oxford University Press, Oxford, 2010).
- ⁹⁹M. A. Nowak and R. M. May, "Evolutionary games and spatial chaos," *Nature* **359**, 826–829 (1992).
- ¹⁰⁰C. Hauert and M. Doebeli, "Spatial structure often inhibits the evolution of cooperation in the snowdrift game," *Nature* **428**, 643–646 (2004).
- ¹⁰¹G. Szabó and C. Tóke, "Evolutionary prisoner's dilemma game on a square lattice," *Phys. Rev. E* **58**, 69–73 (1998).
- ¹⁰²F. C. Santos, M. D. Santos, and J. M. Pacheco, "Social diversity promotes the emergence of cooperation in public goods games," *Nature* **454**, 213–216 (2008).
- ¹⁰³X. Han, Z. Shen, W.-X. Wang, and Z. Di, "Robust reconstruction of complex networks from sparse data," *Phys. Rev. Lett.* **114**, 028701 (2015).
- ¹⁰⁴J. Li, Z. Shen, W.-X. Wang, C. Grebogi, and Y.-C. Lai, "Universal data-based method for reconstructing complex networks with binary-state dynamics," *Phys. Rev. E* **95**, 032303 (2017).
- ¹⁰⁵F. Santosa and W. W. Symes, "Linear inversion of band-limited reflection seismograms," *SIAM J. Sci. Stat. Comp.* **7**, 1307–1330 (1986).
- ¹⁰⁶J. Friedman, T. Hastie, and R. Tibshirani, *The Elements of Statistical Learning* (Springer, Berlin, 2001).
- ¹⁰⁷F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg *et al.*, "Scikit-learn: Machine learning in python," *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
- ¹⁰⁸A. Barrat, M. Barthelemy, and A. Vespignani, *Dynamical Processes on Complex Networks* (Cambridge University Press, 2008).
- ¹⁰⁹A. Kumar, S. Rotter, and A. Aertsen, "Spiking activity propagation in neuronal networks: Reconciling different perspectives on neural coding," *Nat. Rev. Neurosci.* **11**, 615–627 (2010).
- ¹¹⁰G. Szabó and G. Fath, "Evolutionary games on graphs," *Phys. Rep.* **446**, 97–216 (2007).
- ¹¹¹R. Pastor-Satorras, C. Castellano, P. Van Mieghem, and A. Vespignani, "Epidemic processes in complex networks," *Rev. Mod. Phys.* **87**, 925–979 (2015).
- ¹¹²J. Shao, S. Havlin, and H. E. Stanley, "Dynamic opinion model and invasion percolation," *Phys. Rev. Lett.* **103**, 018701 (2009).
- ¹¹³C. Granell, S. Gómez, and A. Arenas, "Dynamical interplay between awareness and epidemic spreading in multiplex networks," *Phys. Rev. Lett.* **111**, 128701 (2013).
- ¹¹⁴F. C. Santos and J. M. Pacheco, "Scale-free networks provide a unifying framework for the emergence of cooperation," *Phys. Rev. Lett.* **95**, 098104 (2005).
- ¹¹⁵A. Koseska, E. Volkov, and J. Kurths, "Oscillation quenching mechanisms: Amplitude vs oscillation death," *Phys. Rep.* **531**, 173–199 (2013).
- ¹¹⁶S. V. Buldyrev, R. Parshani, G. Paul, H. E. Stanley, and S. Havlin, "Catastrophic cascade of failures in interdependent networks," *Nature* **464**, 1025–1028 (2010).
- ¹¹⁷M. Galbiati, D. Delpini, and S. Battiston, "The power to control," *Nat. Phys.* **9**, 126–128 (2013).

- ¹¹⁸D. Balcan and A. Vespignani, "Phase transitions in contagion processes mediated by recurrent mobility patterns," *Nat. Phys.* **7**, 581–586 (2011).
- ¹¹⁹V. Sood and S. Redner, "Voter model on heterogeneous graphs," *Phys. Rev. Lett.* **94**, 178701 (2005).
- ¹²⁰R. Pastor-Satorras and A. Vespignani, "Epidemic spreading in scale-free networks," *Phys. Rev. Lett.* **86**, 3200–3203 (2001).
- ¹²¹C. Castellano, S. Fortunato, and V. Loreto, "Statistical physics of social dynamics," *Rev. Mod. Phys.* **81**, 591 (2009).
- ¹²²A. Bashan, Y. Berezin, S. V. Buldyrev, and S. Havlin, "The extreme vulnerability of interdependent spatially embedded networks," *Nat. Phys.* **9**, 667–672 (2013).
- ¹²³P. L. Krapivsky, S. Redner, and E. Ben-Naim, *A Kinetic View of Statistical Physics* (Cambridge University Press, 2010).
- ¹²⁴J. P. Gleeson, "Binary-state dynamics on complex networks: Pair approximation and beyond," *Phys. Rev. X* **3**, 021004 (2013).
- ¹²⁵S. Rudy, S. L. Brunton, J. L. Proctor, and J. N. Kutz, "Data-driven discovery of partial differential equations," *Sci. Adv.* **3**, e1602614 (2017).
- ¹²⁶X. Li, L. Li, Z. Yue, X. Tang, H. U. Voss, J. Kurths, and Y. Yuan, "Sparse learning of partial differential equations with structured dictionary matrix," *Chaos* **29**, 043130 (2019).
- ¹²⁷P. A. K. Reinbold and R. O. Grigoriev, "Data-driven discovery of partial differential equation models with latent variables," *Phys. Rev. E* **100**, 022219 (2019).
- ¹²⁸D. R. Gurevich, P. A. K. Reinbold, and R. O. Grigoriev, "Robust and optimal sparse regression for nonlinear PDE models," *Chaos* **29**, 103113 (2019).
- ¹²⁹P. A. K. Reinbold, D. R. Gurevich, and R. O. Grigoriev, "Using noisy or incomplete data to discover models of spatiotemporal dynamics," *Phys. Rev. E* **101**, 010203 (2020).
- ¹³⁰B. V. Chirikov and F. M. Izraelev, "Some numerical experiments with a nonlinear mapping: Stochastic component," *Colloques. Int. du CNRS* **229**, 409–428 (1973).
- ¹³¹B. V. Chirikov, "A universal instability of many-dimensional oscillator systems," *Phys. Rep.* **52**, 263–379 (1979).
- ¹³²K. Ikeda, "Multiple-valued stationary state and its instability of the transmitted light by a ring cavity system," *Opt. Commun.* **30**, 257–261 (1979).
- ¹³³S. M. Hammel, C. K. R. T. Jones, and J. V. Moloney, "Global dynamical behavior of the optical field in a ring cavity," *J. Opt. Soc. Am. B* **2**, 552–564 (1985).
- ¹³⁴H. Jaeger, "The 'echo state' approach to analysing and training recurrent neural networks—With an erratum note," GMD Technical Report 148 (German National Research Center for Information Technology, Bonn, Germany), 2001.
- ¹³⁵W. Mass, T. Nachtschlaeger, and H. Markram, "Real-time computing without stable states: A new framework for neural computation based on perturbations," *Neural Comp.* **14**, 2531–2560 (2002).
- ¹³⁶H. Jaeger and H. Haas, "Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication," *Science* **304**, 78–80 (2004).
- ¹³⁷G. Manjunath and H. Jaeger, "Echo state property linked to an input: Exploring a fundamental characteristic of recurrent neural networks," *Neural Comp.* **25**, 671–696 (2013).
- ¹³⁸N. D. Haynes, M. C. Soriano, D. P. Rosin, I. Fischer, and D. J. Gauthier, "Reservoir computing with a single time-delay autonomous Boolean node," *Phys. Rev. E* **91**, 020801 (2015).
- ¹³⁹L. Larger, A. Baylón-Fuentes, R. Martinenghi, V. S. Udaltsov, Y. K. Chembo, and M. Jacquot, "High-speed photonic reservoir computing using a time-delay-based architecture: Million words per second classification," *Phys. Rev. X* **7**, 011015 (2017).
- ¹⁴⁰J. Pathak, Z. Lu, B. Hunt, M. Girvan, and E. Ott, "Using machine learning to replicate chaotic attractors and calculate Lyapunov exponents from data," *Chaos* **27**, 121102 (2017).
- ¹⁴¹Z. Lu, J. Pathak, B. Hunt, M. Girvan, R. Brockett, and E. Ott, "Reservoir observers: Model-free inference of unmeasured variables in chaotic systems," *Chaos* **27**, 041102 (2017).
- ¹⁴²Z. Lu, B. R. Hunt, and E. Ott, "Attractor reconstruction by machine learning," *Chaos* **28**, 061104 (2018).
- ¹⁴³J. Pathak, A. Wilner, R. Fussell, S. Chandra, B. Hunt, M. Girvan, Z. Lu, and E. Ott, "Hybrid forecasting of chaotic processes: Using machine learning in conjunction with a knowledge-based model," *Chaos* **28**, 041101 (2018).
- ¹⁴⁴J. Pathak, B. Hunt, M. Girvan, Z. Lu, and E. Ott, "Model-free prediction of large spatiotemporally chaotic systems from data: A reservoir computing approach," *Phys. Rev. Lett.* **120**, 024102 (2018).
- ¹⁴⁵T. L. Carroll, "Using reservoir computers to distinguish chaotic signals," *Phys. Rev. E* **98**, 052209 (2018).
- ¹⁴⁶K. Nakai and Y. Saiki, "Machine-learning inference of fluid variables from data using reservoir computing," *Phys. Rev. E* **98**, 023111 (2018).
- ¹⁴⁷Z. S. Roland and U. Parlitz, "Observing spatio-temporal dynamics of excitable media using reservoir computing," *Chaos* **28**, 043118 (2018).
- ¹⁴⁸T. Weng, H. Yang, C. Gu, J. Zhang, and M. Small, "Synchronization of chaotic systems and their machine-learning models," *Phys. Rev. E* **99**, 042203 (2019).
- ¹⁴⁹A. Griffith, A. Pomerance, and D. J. Gauthier, "Forecasting chaotic systems with very low connectivity reservoir computers," *Chaos* **29**, 123108 (2019).
- ¹⁵⁰J. Jiang and Y.-C. Lai, "Model-free prediction of spatiotemporal dynamical systems with recurrent neural networks: Role of network spectral radius," *Phys. Rev. Res.* **1**, 033056 (2019).
- ¹⁵¹P. R. Vlachas, J. Pathak, B. R. Hunt, T. P. Sapsis, M. Girvan, E. Ott, and P. Koumoutsakos, "Forecasting of spatio-temporal chaotic dynamics with recurrent neural networks: A comparative study of reservoir computing and backpropagation algorithms," [arXiv:1910.05266](https://arxiv.org/abs/1910.05266) (2019).
- ¹⁵²H. Fan, J. Jiang, C. Zhang, X. Wang, and Y.-C. Lai, "Long-term prediction of chaotic systems with machine learning," *Phys. Rev. Res.* **2**, 012080 (2020).
- ¹⁵³C. Zhang, J. Jiang, S.-X. Qu, and Y.-C. Lai, "Predicting phase and sensing phase coherence in chaotic systems with machine learning," *Chaos* **30**, 083114 (2020).
- ¹⁵⁴L.-W. Kong, H.-W. Fan, C. Grebogi, and Y.-C. Lai, "Machine learning prediction of critical transition and system collapse," *Phys. Rev. Res.* **3**, 013090 (2021).
- ¹⁵⁵L.-W. Kong, H.-W. Fan, C. Grebogi, and Y.-C. Lai, "Emergence of transient chaos and intermittency in machine learning," *J. Phys. Complex.* **2**, 035014 (2021).