

## EXACT RECOVERY OF CHAOTIC SYSTEMS FROM HIGHLY CORRUPTED DATA\*

GIANG TRAN<sup>†</sup> AND RACHEL WARD<sup>†</sup>

**Abstract.** Learning the governing equations in dynamical systems from time-varying measurements is of great interest across different scientific fields. This task becomes prohibitive when such data is, moreover, highly corrupted, for example, due to the recording mechanism failing over unknown intervals of time. When the underlying system exhibits chaotic behavior, such as sensitivity to initial conditions, it is crucial to recover the governing equations *with high precision*. In this work, we consider continuous time dynamical systems  $\dot{x} = f(x)$  where each component of  $f : \mathbb{R}^d \rightarrow \mathbb{R}^d$  is a multivariate polynomial of maximal degree  $p$ ; we aim to identify  $f$  exactly from possibly highly corrupted measurements  $x(t_1), x(t_2), \dots, x(t_m)$ . As our main theoretical result, we show that if the system is sufficiently ergodic that this data satisfies a strong central limit theorem (as is known to hold for chaotic Lorenz systems), then the governing equations  $f$  can be *exactly recovered* as the solution to an  $\ell_1$  minimization problem—even if a large percentage of the data is corrupted by outliers. Numerically, we apply the alternating minimization method to solve the corresponding constrained optimization problem. Through several examples of three-dimensional chaotic systems and higher-dimensional hyperchaotic systems, we illustrate the power, generality, and efficiency of the algorithm for recovering governing equations from noisy and highly corrupted measurement data.

**Key words.** dynamical systems, system identification, compressed sensing, sparse representation,  $\ell_1$ -minimization, splitting optimization methods, chaotic systems, ergodic theory

**AMS subject classifications.** 37E99, 65K10, 65P20, 34H10

**DOI.** 10.1137/16M1086637

**1. Introduction.** Discovering the underlying dynamical equations from time-dependent observations is of great interest across many scientific fields. Examples include statistical learning theory [52], manifold learning [41], machine learning [22], physical modeling [44], and system identification [27, 28]. In those examples, it is generally assumed that the governing equations can be expressed as a combination of terms in an appropriate functional space [21]. The selection of the important terms within this space is obtained from regularization, pruning, shrinking, or regression. An overview of reconstruction methods for dynamical systems can be found in [46]. Indeed, it is shown that without additional information, learning the governing equations from measurement data is *intractable* [15], and may face the *curse of dimensionality* [19], regardless of how much data there is.

Identification of nonlinear dynamical systems is one of the most active areas in system identification and one of the main topics in the rapid development of nonlinear dynamics [1, 28]. It is shown that simple nonlinearities in the governing equations can lead to incredibly complicated behavior in the solutions of the nonlinear systems, which is the so-called *butterfly effect* present in chaotic solutions, [47, 48, 30]. For example, the well-known three-variable Lorenz system [29] has seven terms with only one nonlinear term of quadratic type on the right-hand side, and the Rössler system [40] has six terms with only one nonlinear term, yet their solutions exhibit chaotic

\*Received by the editors July 26, 2016; accepted for publication (in revised form) March 13, 2017; published electronically July 25, 2017.

<http://www.siam.org/journals/mms/15-3/M108663.html>

**Funding:** The second author was partially supported by NSF CAREER grant 1255631, AFOSR YIP grant FA9550-13-1-0125, and Fall 2015 Moncrief Grand Challenge Faculty Award.

<sup>†</sup>Department of Mathematics, The University of Texas at Austin, Austin, TX 78712 (gtran@math.utexas.edu, rward@math.utexas.edu).

behavior.

There have been many approaches to extract the underlying structures of chaotic systems from time-dependent data. A review of major methods in modeling nonlinear dynamics and chaotic systems can be found in [1]. One of the main directions is the reconstruction of state space from one-dimensional realizations, [34, 14, 11, 23, 25]. Those methods rely on the estimation of the time delay and the embedding dimension to reconstruct a state space which preserves the topological properties of the original system. In [14, 42], the authors use the singular value decomposition to determine the appropriate dependent variables that will appear in the dynamical equations. The reconstruction guarantee in the absence of noise, called the *delay embedding theorem*, is proved by [49]. However, it is showed that in some special choices of parameters, two different systems can produce the exact same time series of one of their variables [26].

Even with the availability of state-space data, it is challenging to recover the parameters in the governing equations. In [45], the authors use symbolic regression to find both the parameters and the forms of the equations simultaneously. In [7], the authors recast the problem of recovering coefficients in a known basis as a linear regression, where the matrix for regression is built from the data. They, moreover, incorporate thresholding in the regression to further promote sparsity in the recovered coefficients that govern the chaotic systems.

Sparsity has been playing a significant role in the developments of compressed sensing, image processing, optimization, and many others. Recently, sparse-inducing methods used in image processing and compressed sensing have been applied to partial differential equations, dynamical systems, and physical sciences [54, 43, 12, 6, 32, 33, 37, 31, 8, 9, 50, 7]. In these works, the authors study either the sparse property of the solutions in different contexts or the sparse structures of the governing equations. In the latter direction, it turns out that many chaotic systems have simple algebraic representations corresponding to a sparse representation in high-dimensional nonlinear functional spaces [47, 48, 30]. In [35, 36], the authors provide reconstruction guarantee for linear stochastic differential equation using mutual information.

In this work, we bring together connections between compressed sensing, splitting optimization methods, sparse representations of the governing equations, and the statistical properties of chaotic systems, to provide *exact recovery guarantees* for classes of chaotic systems. In particular, we provide conditions and algorithms for recovering the governing equations from possibly highly corrupted data generated from a class of *Lorenz-like systems* which includes the classical Lorenz equations, and are known to be ergodic. Explicitly, *when the underlying attractor of the flow has Hausdorff dimension greater than two, the flow satisfies some mixing properties and the governing equation vector  $f$  has a sparse representation in the space of multivariable polynomials, then the polynomial coefficients of  $f$  as well as the outlier vectors can be exactly recovered as the unique solution to a partial  $\ell_1$ -minimization problem with high probability (depends on the number of measurements), as long as the number of measurements is big enough and sparse level is low enough.* We prove theoretical reconstruction guarantees by combining the partial sparse recovery results [5] with statistical behavior of the Lorenz-like systems [2, 4]. It is based on the observation that although individual trajectories are highly unpredictable due to the sensitivity property to initial conditions of chaotic dynamic systems, their statistical behavior is understandable and share many of the same properties of random sequences. We finally mention that for our theoretical results, we do not necessarily require the system to be chaotic in the formal sense, but rather we use certain ergodicity properties

which are satisfied for a class of geometric Lorenz attractors which are also known to be chaotic. Therefore, our theory can likely be generalized to other dynamical systems with similar ergodicity properties.

The paper is divided as follows. In section 2, we explain the problem setting. In section 3, we first review some results from compressed sensing and statistical properties of chaotic systems. Then we present our theoretical reconstruction guarantees and state in which conditions the sparse solutions can be recovered. The numerical implementations and results are described in sections 4 and 5. Concluding remarks and discussion are given in section 6.

**2. Problem setting.** In this work, we are interested in reconstructing the governing equations of a chaotic system from the time-varying measurement data in which data is corrupted at unknown intervals of time. Here, we consider chaotic systems of the form

$$(1) \quad \frac{d}{dt}x(t) = f(x(t)),$$

where the column vector  $x(t) = (x_1(t), x_2(t), \dots, x_d(t))^T$  represents the state of the system at time  $t$  and the nonlinear function vector  $f(x) = (f_1(x), f_2(x), \dots, f_d(x))^T$  defines the dynamic motions. Even though chaotic systems are deterministic, they exhibit stochastic behavior [29]. Such behavior is commonly referred to as the *butterfly effect*, i.e., small differences in initial conditions will yield much larger differences in the outcomes. Therefore, it is very important to recover the governing equations with high accuracy.

In many chaotic systems inspired by physical and biological processes, the governing equations may consist of only a few terms in a high-dimensional nonlinear functional space [29, 40, 47, 48, 30]. This observation also holds for many dynamical systems and partial differential equations where the complicated system can be well-approximated by a simpler system with similar behavior [43, 7]. Explicitly, for many chaotic systems, the governing equations can be represented in the space of polynomial functions

$$(2) \quad f_j(x(t)) = c_{j0} + \sum_k c_{j,k} x_k(t) + \sum_{k,l} c_{j,k,l} x_k(t) x_l(t) + \sum_{k,l,n} c_{j,k,l,n} x_k(t) x_l(t) x_n(t) + \dots,$$

where the coefficient vectors  $c_j = (c_{j,0}, c_{j,1}, \dots, c_{j,d}, c_{j,1,1}, \dots, c_{j,d,d}, \dots)$  are, moreover, sparse. Note that  $c_j$  is a vector of length

$$(3) \quad r = \binom{p+d}{d} \leq \left( \frac{p+d}{d} \right)^d,$$

the maximal number of monomials of degree at most  $p$  in a multivariate polynomial in  $d$  variables. In fact, many key dynamical systems arising from applications in biology and physics, such as the Lorenz system, are in fact *bilinear*, so that  $p = 2$ .

For the rest of the paper, we denote  $d$  to be the dimension of the system,  $m$  the number of measurements, and  $r$  the cardinal of the basis for a given nonlinear functional space. Given a collection of data at different times  $\{x(t_1), x(t_2), \dots, x(t_m)\}$ , as in [7], we construct the following matrices  $\tilde{X}$ ,  $\dot{\tilde{X}}$ , and  $\Phi(\tilde{X})$  to store the measurement data, the time derivative of the data, and the dictionary built from the data:

$$(4) \quad X = \begin{bmatrix} | & | & & | \\ X_1 & X_2 & \dots & X_d \\ | & | & & | \end{bmatrix} = \begin{bmatrix} x_1(t_1) & x_2(t_1) & \dots & x_d(t_1) \\ x_1(t_2) & x_2(t_2) & \dots & x_d(t_2) \\ \vdots & \vdots & \dots & \vdots \\ x_1(t_m) & x_2(t_m) & \dots & x_d(t_m) \end{bmatrix}_{m \times d},$$

$$(5) \quad \dot{X} = \begin{bmatrix} | & | & & | \\ \dot{X}_1 & \dot{X}_2 & \dots & \dot{X}_d \\ | & | & & | \end{bmatrix}_{m \times d} = \begin{bmatrix} \dot{x}_1(t_1) & \dot{x}_2(t_1) & \dots & \dot{x}_d(t_1) \\ \dot{x}_1(t_2) & \dot{x}_2(t_2) & \dots & \dot{x}_d(t_2) \\ \vdots & \vdots & \dots & \vdots \\ \dot{x}_1(t_m) & \dot{x}_2(t_m) & \dots & \dot{x}_d(t_m) \end{bmatrix}_{m \times d},$$

$$(6) \quad \Phi(X) = \begin{bmatrix} | & | & | & | & \\ 1 & X & X^{P_2} & X^{P_3} & \dots \\ | & | & | & | & \end{bmatrix}_{m \times r},$$

where for each positive integer  $k$ ,  $X^{P_k}$  denotes the values of all monomials of total degree  $k$  at the times  $t_1, t_2, \dots, t_m$ . For example,  $X^{P_2}$  is a matrix of size  $m \times \left(\frac{d(d+1)}{2}\right)$

$$X^{P_2} = \begin{bmatrix} x_1^2(t_1) & x_1(t_1)x_2(t_1) & \dots & x_1(t_1)x_d(t_1) & x_2^2(t_1) & \dots & x_d^2(t_1) \\ x_1^2(t_2) & x_1(t_2)x_2(t_2) & \dots & x_1(t_2)x_d(t_2) & x_2^2(t_2) & \dots & x_d^2(t_2) \\ \vdots & \vdots & \dots & \vdots & \dots & \vdots & \vdots \\ x_1^2(t_m) & x_1(t_m)x_2(t_m) & \dots & x_1(t_m)x_d(t_m) & x_2^2(t_m) & \dots & x_d^2(t_m) \end{bmatrix}.$$

In the case of nondamaged data, [7] observed that the problem of finding  $f$  of the form (2) in the case of exact data  $X$  and  $\dot{X}$  can be reformulated as finding the coefficient matrix  $\mathcal{C} = [c_1 \ c_2 \ \dots \ c_d]_{r \times d}$  such that  $\dot{X} = \Phi(X)\mathcal{C}$ . Thus, given noisy data, such as if  $\dot{X}$  is approximated via finite differences of successive values of  $X$ , they propose to solve for  $\mathcal{C}$  via linear regression:

$$(7) \quad \mathcal{C} = \underset{\tilde{\mathcal{C}}}{\operatorname{argmin}} \frac{1}{2} \|\dot{X} - \Phi(X)\tilde{\mathcal{C}}\|_2^2.$$

The authors, moreover, incorporate hard thresholding into the regression procedure to promote sparsity in  $\mathcal{C}$ . They illustrate the effectiveness of their algorithm for stably reconstructing the governing coefficients in the presence of a small amount of noise, but do not provide theoretical performance guarantees. In particular, there are no guarantees a priori that the matrix  $\Phi(X)$  will be full rank, which is necessary for uniqueness of the recovered  $\mathcal{C}$ . For example, the matrix  $\Phi(X)$  will not have full rank if  $x(t_k) = x(t_{k+1})$  is at a fixed point, or if  $x(t_{k+L}) = x(t_k)$  is fixed at a cycle of length  $L < r$  more generally.

As a by-product of our main theorem, we provide conditions under which the regression algorithm from [7] is theoretically justified; namely, will show that *for chaotic data, the columns are independent, and thus the matrix  $\Phi(X)$  has full column rank and thus the polynomial coefficients generating the data is unique*. See Corollary 3.13 for more details. More generally, we are interested in the case when the data might be highly corrupted, such as over unknown intervals of time due to disruption of the measurement device.

Our set-up is as follows: consider data  $X^\circ$  which is corrupted in the sense that some small fraction of the  $m$  measurements  $\{x(t_1), x(t_2), \dots, x(t_m)\}$  and their time derivatives  $\{\dot{x}(t_1), \dot{x}(t_2), \dots, \dot{x}(t_m)\}$  are perturbed by bounded additive error:

$x^o(t_j) = x(t_j) + \theta_j$  and  $\dot{x}^o(t_j) = \dot{x}(t_j) + \theta'_j$ . Exploiting that the corruptions are sparse, the matrix difference  $\mathcal{E} = \dot{X}^o - \Phi(X^o)\mathcal{C}$  will only have a small fraction of its rows which are nonzero. Thus, denoting the  $j$ th row of  $\mathcal{E}$  by  $\mathcal{E}(j, :)$ , the optimization problem of jointly recovering the polynomial governing coefficients and locations of the corruptions can be posed as a sparse recovery problem:

$$(8) \quad \begin{aligned} \min_{(\mathcal{C}, \mathcal{E})} \quad & \|\{j : \mathcal{E}(j, :) \text{ is nonzero}\}\| \\ \text{subject to (s.t.)} \quad & \Phi(X^o)\mathcal{C} + \mathcal{E} = \dot{X}^o. \end{aligned}$$

Of course, this optimization problem is intractable, so we relax the objective function to be convex. To enforce the group-sparsity with respect to the rows of  $\mathcal{E}$  [18, 24, 16], we consider the following optimization model

$$(9) \quad \begin{aligned} \min_{(\mathcal{C}, \mathcal{E})} \quad & \|\mathcal{E}\|_{2,1} = \min_{(\mathcal{C}, \mathcal{E})} \sum_{j=1}^m \|\mathcal{E}(j, :)\|_2 \\ \text{s.t.} \quad & \Phi(X^o)\mathcal{C} + \mathcal{E} = \dot{X}^o. \end{aligned}$$

For more details, see section 3. To enforce additional sparsity in the polynomial coefficient matrix  $\mathcal{C}$ , we will also consider the problem

$$(10) \quad \begin{aligned} \min_{(\mathcal{C}, \mathcal{E})} \quad & \|\mathcal{E}\|_{2,1} = \min_{(\mathcal{C}, \mathcal{E})} \sum_{j=1}^m \|\mathcal{E}(j, :)\|_2 \\ \text{s.t.} \quad & \Phi(X^o)\mathcal{C} + \mathcal{E} = \dot{X}^o \quad \text{and} \quad \mathcal{C} \text{ is sparse.} \end{aligned}$$

For more details, see sections 4 and 5.

**3. Reconstruction guarantee analysis.** In this section, we first recall some results from compressive sensing and, in particular, partial sparse recovery problems therein, as well as statistical properties of Lorenz-like systems. Then we present our theoretical guarantees for the framework (9).

**3.1. Theory from compressive sensing.** The compressive sensing paradigm, in its most basic form as introduced in [17, 10], considers, for an underdetermined linear system of equations  $y = Ax$ , the NP hard minimization problem

$$(11) \quad \begin{aligned} x_0 = \operatorname{argmin}_z \quad & \|z\|_0 = |\{j : |z_j| > 0\}| \\ \text{s.t.} \quad & Az = y. \end{aligned}$$

The convex relaxation of this problem is the  $\ell_1$  minimization problem

$$(12) \quad \begin{aligned} x_1 = \operatorname{argmin}_z \quad & \|z\|_1 = \sum_{j=1}^n |z_j| \\ \text{s.t.} \quad & Az = y. \end{aligned}$$

Compressive sensing theory provides conditions on the underdetermined matrix  $A$  such that the solutions  $x_0$  and  $x_1$  are equivalent and equal to  $x$  satisfying  $Ax = y$  whenever there exists such an  $x$  is sufficiently sparse. The theoretical guarantees are also stable with respect to nonexact sparse solutions and robust with respect to additive noise on the measurements  $y = Ax$ , but for simplicity we discuss only the

case of exact sparsity here. For a comprehensive overview of compressive sensing, we refer the reader to [20].

From here on out, we say that a vector  $x \in \mathbb{R}^n$  is  $s$ -sparse. As shown in [17, 10, 13], a certain *null-space property* for an  $m \times n$  matrix  $A$  is a sufficient and necessary condition for sparse solutions to be exactly recovered via  $\ell_1$  minimization.

**PROPOSITION 3.1.** *Given a matrix  $A \in \mathbb{R}^{m \times n}$ , every  $s$ -sparse vector  $x$  is the unique solution of (12) with  $y = Ax$  if and only if for every  $v \in \mathbb{R}^n \setminus \{0\}$  in the null space of  $A$ , and for every set  $S \subset \{1, 2, \dots, n\}$  of cardinality  $s$ , the following holds:*

$$\|v_S\|_1 < \frac{1}{2}\|v\|_2.$$

In other words, the null space property means that all vectors in the null space of the measurement matrix  $A$  should be sufficiently unconcentrated on any subset of its entries, or rather, sufficiently “far” from the nonlinear set of sparse vectors.

More recently, a variant of compressive sensing theory has been developed for the theory of *partial sparse* recovery. Although there is rich literature on partial sparse recovery guarantees, we will use the results from [5] which are most closely related to the setting at hand.

We use the following definition and theorem for *partial sparse recovery* from [5].

**DEFINITION 3.2** (partial null space property). *A pair of matrices  $(A_1^{l \times (n-r)}, A_2^{l \times r})$  satisfies the null space property (NSP) of order  $s - r$  for partially sparse recovery of size  $n - r$  with  $r \leq s$  if  $A_2$  is full column rank and if for every  $v_1 \in \mathbb{R}^{n-r} \setminus \{0\}$  such that  $A_1 v_1 \in \mathcal{R}(A_2)$ , the range of  $A_2$ , and for every set  $S \subset \{1, \dots, n-r\}$  of cardinality  $s - r$ , the following holds:*

$$\|(v_1)_S\|_1 < \frac{1}{2}\|A_1 v_1\|_1.$$

**PROPOSITION 3.3.** *A pair of matrices  $A = (A_1^{l \times (n-r)}, A_2^{l \times r})$  satisfies the NSP of order  $s - r$  for partially sparse recovery of size  $n - r$  if and only if every  $\bar{v} = (\bar{v}_1, \bar{v}_2)$ , such that  $\bar{v}_1 \in \mathbb{R}^{n-r}$  is  $(s - r)$ -sparse and  $\bar{v}_2 \in \mathbb{R}^r$  such that  $A_1 \bar{v}_1 + A_2 \bar{v}_2 = y$ , is the unique solution to*

$$\min_{v'_1, v'_2} \|v'_1\|_1 \text{ s.t. } A_1 v'_1 + A_2 v'_2 = y.$$

We will use this proposition in the particular case  $A_1 = Id_{l \times l}$ . We state this special case explicitly for clarity.

**COROLLARY 3.4.** *Every  $(v, w) \in \mathbb{R}^{m+r}$  satisfying  $v + Aw = y$  such that  $v \in \mathbb{R}^m$  is  $(s - r)$ -sparse is the unique solution to*

$$\min_{\tilde{v}, \tilde{w}} \|\tilde{v}\|_1 \text{ s.t. } \tilde{v} + A\tilde{w} = y,$$

*if and only if  $A \in \mathbb{R}^{m \times r}$  is full column rank and for every  $\tilde{v} \in \mathbb{R}^m \setminus \{0\}$  such that  $\tilde{v} \in \mathcal{R}(A)$ , the following holds for every set  $S \subset \{1, \dots, m\}$  of cardinality  $s - r$ :*

$$\|\tilde{v}_S\|_1 < \frac{1}{2}\|\tilde{v}\|_1.$$

Note that a straightforward corollary of this corollary is that exact and unique recovery still holds if we add any additional consistent linear constraints to the linear program.

COROLLARY 3.5. *Under the same conditions as above, if for some pair of matrices  $(B_1, B_2)$  the solution vector  $(v, w)$  satisfies additionally  $B_1v + B_2w = z$ , then under the same conditions  $(v, w)$  is the unique minimizer to the program*

$$\min_{\tilde{v}, \tilde{w}} \|\tilde{v}\|_1 \text{ s.t. } \tilde{v} + A\tilde{w} = y, \quad B_1\tilde{v} + B_2\tilde{w} = z.$$

**3.2. Statistical behavior of Lorenz-like systems.** In this section, we are interested in three-dimensional (3D) dynamical systems, i.e., the dimension of the system is  $d = 3$ . We first recall the following notation. We denote by  $C^{1+\eta}(\Omega)$  the Holder space consisting of those functions  $f$  having continuous derivative up to order 1 and such that all partial derivatives are Holder continuous with exponent  $\eta$ :

$$\|f\|_{C^{1+\eta}(\Omega)} := \max_{|\beta| \in \{0,1\}} \sup_{x \in \Omega} |D^\beta f(x)| + \max_{|\beta|=1} \sup_{x \neq y \in \Omega} \frac{|D^\beta f(x) - D^\beta f(y)|}{|x - y|^\eta},$$

where  $\beta$  ranges over multi-indices with  $|\beta| = \sum_i |\beta_i|$ .

Recall the well-known classical *Lorenz equations*:

$$(13) \quad \begin{aligned} \dot{x}_1 &= a(x_2 - x_1), & a &= 10, \\ \dot{x}_2 &= \gamma x_1 - x_2 - x_1 x_3, & \gamma &= 28, \\ \dot{x}_3 &= x_1 x_2 - b x_3, & b &= 8/3. \end{aligned}$$

Note that the Lorenz equations are of the form  $\dot{x}(t) = f(x(t))$  with polynomial governing equations  $f$  of the form (2) of degree  $p = 2$ . With slight abuse of notation, we will go back and forth also between the notation  $\dot{X}^t = f(X^t)$ .

The author of [29] introduced these equations as a simplified model for weather forecast, and numerical simulations indicated that in an open neighborhood of the chosen parameters, almost all points in phase space tend to a chaotic attractor. One property of chaotic systems is *sensitive dependence on initial conditions*, which implies that long term predictions based on such models are infeasible; on the other hand, another property of chaos is that the *statistical* behavior of such systems is understandable, and chaotic systems share many of the same statistical properties of random sequences. The Lorenz attractor, while easy to visualize numerically, has proved extremely difficult to analyze rigorously. A proof of existence of the Lorenz attractor was only provided in 1999 [51], incorporating a computer-aided proof. Precisely, Tucker proved that the Lorenz equations (13) support a compact, connected attractor  $\Lambda$  and the flow admits a unique so-called “physical” measure  $\mu$  with  $\text{supp}(\mu) = \Lambda$ . An invariant probability measure  $\mu$  for a flow  $X^t := (x_1(t), x_2(t), x_3(t))$  on a compact Riemannian manifold  $M$  is called physical if the *basin* of  $\mu$ ,  $B(\mu)$  has positive Lebesgue measure. Recall that  $B(\mu)$  is the set of points  $z \in M$  satisfying for all continuous functions  $\psi : M \rightarrow \mathbb{R}$

$$(14) \quad \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \psi(X^t(z)) dt = \int_\Lambda \psi(z) d\mu(z).$$

Roughly speaking, the existence of a physical measure for an attractor means that most points in a neighborhood of the attractor have well-defined long term statistical behavior. For the Lorenz equations (13), and more generally, for any so-called *geometric Lorenz-like system* (see [2] for definition and properties) of a flow on a 3D manifold, the ergodic basin  $B(\mu)$  covers a full Lebesgue measure subset of the topological basin of attraction  $\Lambda$ .

Property (14) of a physical measure shows that, asymptotically, the time average of a continuous observable of the flow equals its space average. It is natural to ask more quantitatively for the *rate of convergence* of the time averages to the space average and, moreover, if such ergodicity also holds for the time-1 map for the flow, given by the discrete sequence  $\{X^j\}_{j \in \mathbb{Z}}$ . Following several results in this direction for certain classes of so-called geometric Lorenz attractors which include the Lorenz equations (13), we state one of the most recent results in this direction. Combining Theorems 5.2 and 7.1 from [4] with Corollary 2.2 of [3], one can derive estimates on the rate of mixing in (14) for the discrete time-1 map in the form of an *almost sure invariance principle* (ASIP) for the time-1 map of flows generated by the Lorenz equations (13) and, more generally, for the time-1 map of a vector field belonging to the following (quite technical) class.

**DEFINITION 3.6.** Denote by  $\mathcal{U}$  the class of  $C^{1+\eta}$  uniformly hyperbolic skew product flows, subject to a uniform nonintegrability condition, as defined by the condition UNI and properties (i)–(iv) from [3]. This class includes the classical Lorenz attractor, and an open set of coefficients around the classical coefficients.

**PROPOSITION 3.7** (ASIP for time-1 maps). Fix  $\eta > 0$ . Let  $X^t$  be the flow generated by a vector field  $G \in \mathcal{U}$  starting from  $X^0 = x \in \Lambda$ , and consider its time-1 map  $\{X^0, X^1, X^2, \dots, X^m\}$ . Let  $\psi: \mathbb{R}^3 \rightarrow \mathbb{R}$  be a  $C^{1+\eta}$  function. Then there is a standard normal random variable  $Z$ , a universal constant  $C_1 > 0$ , and a constant  $C_{x,\psi} \geq 0$  such that

$$\left| \frac{1}{m} \sum_{j=0}^{m-1} \psi(X^j) - \int_{\Lambda} \psi(z) d\mu(z) - \frac{\sigma}{\sqrt{m}} Z \right| \leq C_{x,\psi} m^{-3/4} (\log(m))^{1/2} (\log \log(m))^{1/4},$$

for  $\mu$ -almost all  $x \in \Lambda$ ,

and the variance is bounded by  $\sigma^2 \leq C_1 \|\psi\|_{C^{1+\eta}(\Lambda)}^2$ .

**Remark 3.8.** Note that the ASIP implies the central limit theorem (CLT): in the same setting as above,

$$\frac{1}{\sqrt{m}} \left( \sum_{j=0}^{m-1} \psi(X^j) - m \int_{\Lambda} \psi(x) d\mu(x) \right) \rightarrow \mathcal{N}(0, \sigma^2) \quad \text{as } m \rightarrow \infty,$$

where the convergence is in distribution.

For our purposes, it will be useful to state a more uniform version of Proposition 3.7. Consider the function

$$(15) \quad F_{\tau,\eta}(x) = \sup_{\psi: \|\psi\|_{C^{1+\eta}} \leq \tau} C_{x,\psi}.$$

By Proposition 3.7,  $F_{\tau}(x)$  is finite for  $\mu$ -almost all  $x$ . Thus, given any  $\varepsilon > 0$ , there exists a constant  $\kappa_{\varepsilon,\tau,\eta}$  and a subset  $\Lambda_{\varepsilon,\tau,\eta} \subset \Lambda$  of measure  $\mu(\Lambda_{\varepsilon,\tau,\eta}) \geq 1 - \varepsilon$  such that  $F_{\tau,\eta}(x) \leq \kappa_{\varepsilon,\tau,\eta}$  uniformly for all  $x \in \Lambda_{\varepsilon,\tau,\eta}$ . In terms of this constant, we can state the following corollary.

**COROLLARY 3.9.** Fix  $\eta > 0$  and  $\varepsilon > 0$ . Let  $X^t$  be the flow generated by a vector field  $G \in \mathcal{U}$ . Draw  $x \in \Lambda$  from the measure  $d\mu$ , and consider the flow  $X^t$  generated by



such a vector field originating at  $X^0 = x$ , and its time-1 map  $\{X^0, X^1, X^2, \dots\}$ . There is a universal constant  $C_1 > 0$  and a constant  $\kappa_{\varepsilon, \tau, \eta} \geq 0$  such that with probability exceeding  $1 - \varepsilon$  with respect to the draw of  $x$ , the following holds uniformly over all  $\psi : \mathbb{R}^3 \rightarrow \mathbb{R}$  satisfying  $\|\psi\|_{C^{1+\eta}} \leq \tau$ :

$$\left| \frac{1}{m} \sum_{j=0}^{m-1} \psi(X^j) - \int_{\Lambda} \psi(z) d\mu(z) - \frac{\sigma}{\sqrt{m}} Z \right| \leq \kappa_{\varepsilon, \tau, \eta} m^{-3/4} (\log(m))^{1/2} (\log \log(m))^{1/4},$$

and the variance is bounded by  $\sigma^2 \leq C_1 \tau^2$ .

For any  $G \in \mathcal{U}$ , the corresponding attractor  $\Lambda$  will be compact, i.e., there exists some finite  $B_{\Lambda} > 0$  such that

$$\max_{(x, y, z) \in \Lambda} \{|x|, |y|, |z|\} \leq B_{\Lambda}.$$

For the Lorenz equations (13) in particular,  $\Lambda$  is a fractal, and has measured Hausdorff dimension  $2.06 \pm .01$  [53].

**3.3. Recovering polynomial dynamics from chaotic data.** Fix number of measurements  $m$ . Fix  $\Theta$  and  $\Theta' \in \mathbb{R}^{m \times 3}$ , arrays of sparse corruptions such that  $\|\Theta(\cdot, l)\|_0, \|\Theta'(\cdot, l)\|_0 \leq s$ , and are uniformly bounded,

$$(16) \quad \sup_j \{\|\Theta(j, l)\|_{\infty}, \|\Theta'(j, l)\|_{\infty}\} \leq B_{\Theta}, \quad l = 1, 2, 3.$$

Suppose that we observe corrupted iterations of the time-1 map of a flow  $X^t = (x_1(t), x_2(t), x_3(t))$  satisfying the conditions of Corollary 3.9:

$$(17) \quad \text{Given: } U^t = X^t + \Theta_t, \quad V^t = \dot{X}^t + \Theta'_t, \quad t = 0, 1, 2, \dots$$

Our measurements  $U^j = (u_1(j), u_2(j), u_3(j))$  and  $V^j = (v_1(j), v_2(j), v_3(j))$  satisfy the linear equations

$$(18) \quad V - \Phi C = \mathcal{E},$$

where

- $\Phi$  has rows  $\Phi(j, \cdot) = (1, u_1(j), u_2(j), u_3(j), u_1(j)u_2(j), \dots)$ ,
- $\mathcal{E} = (e_1, e_2, e_3) \in \mathbb{R}^{m \times 3}$  has a sparse number of nonzero rows, with  $\|\mathcal{E}\|_0 \leq 6s$ ,
- $C = (c_1, c_2, c_3) \in \mathbb{R}^{r \times 3}$  is the matrix of polynomial coefficients.

Notice that by turning the matrices  $V, C, \mathcal{E}$  into tall column vectors  $v, c, e$  and turning  $\Phi$  into the augmented matrix

$$A = \begin{bmatrix} \Phi & 0 & 0 \\ 0 & \Phi & 0 \\ 0 & 0 & \Phi \end{bmatrix},$$

we can equivalently write (18) as  $v - Ac = e$ . The range space  $\mathcal{R}(A)$  corresponds to vectors  $Ac = (\Phi c_1, \Phi c_2, \Phi c_3) \in \mathbb{R}^{3m \times 3}$  of the form

$$(19) \quad \begin{aligned} (\Phi c_l)_j &= \sum_{\alpha: |\alpha| \leq p} c_l(\alpha) u_1(j)^{\alpha_1} u_2(j)^{\alpha_2} u_3(j)^{\alpha_3} \\ &= \sum_{\alpha: |\alpha| \leq p} c_l(\alpha) x_1(j)^{\alpha_1} x_2(j)^{\alpha_2} x_3(j)^{\alpha_3} + R_l(j), \quad l = 1, 2, 3, \end{aligned}$$

where  $\|R_l\|_0 \leq 2s$  and, in light of the assumption (16), is bounded by

$$(20) \quad |R_l(j)| \leq (B_\Theta + B_\Lambda)^d \|c_l\|_1, \quad l = 1, 2, 3.$$

**LEMMA 3.10.** *Let  $\eta \geq 0$ . Suppose that the underlying attractor  $\Lambda$  for the flow at hand has Hausdorff dimension greater than two, and consider the function  $\psi^c = \psi^{(c_1, c_2, c_3)} : \mathbb{R}^3 \rightarrow \mathbb{R}$  given by*

$$\psi^c(x) = \sum_{l=1}^3 \left| \sum_{\alpha: |\alpha| \leq p} c_l(\alpha) x_1^{\alpha_1} x_2^{\alpha_2} x_3^{\alpha_3} \right|^{1+\eta}.$$

Then

$$\inf_{c \in \mathbb{R}^{3r}: \|c\|_1=1} \int_{\Lambda} \psi^c(x) d\mu(x) \geq D > 0.$$

*Proof.* Since  $\Lambda$  has Hausdorff dimension strictly greater than two, it is not contained in the zero set of any algebraic polynomial. Thus, for any fixed nonzero  $c \in \mathbb{R}^{3r}$ ,

$$G(c) := \int_{\Lambda} \sum_{j=1}^m |(\psi^c(x))_j|^{1+\eta} d\mu(x) > 0.$$

Since the set  $c \in \mathbb{R}^{3r} : \|c\|_1 = 1$  is compact and nonempty, we may apply the extreme value theorem: any continuous real-valued function over the space is bounded below and attains its infimum. In particular, this implies

$$\inf_{c \in \mathbb{R}^{3r}: \|c\|_1=1} G(c) > 0.$$

**THEOREM 3.11** (main theorem). *Fix  $\eta > 0, \varepsilon > 0$ , and maximal degree  $p$ . Let  $X^t = x(t) = (x_1(t), x_2(t), x_3(t))$  be the flow generated by a vector field  $G \in \mathcal{U}$  whose governing equation  $f : \mathbb{R}^3 \rightarrow \mathbb{R}^3$  in  $\dot{x}(t) = f(x(t))$  is a multivariate algebraic polynomial of degree at most  $p$ , and suppose we observe corrupted measurements of the time-1 map*

$$U^t = X^t + \Theta_t, \quad V^t = \dot{X}^t + \Theta'_t, \quad t = 0, 1, 2, \dots, m,$$

where  $(\Theta, \Theta') \in \mathbb{R}^{6m}$  is sparse such that  $\|(\Theta, \Theta')\|_0 \leq 2s$ , and the sparse level  $s$  is greater or equal than  $\binom{p+d}{d}$ , the maximal number of monomials of degree at most  $p$ . Also, assume that the underlying attractor for the flow has Hausdorff dimension greater than two.

Then there are constants  $C, C'$  depending only on  $\Lambda, p, B_\Theta, \eta$ , and  $\varepsilon$  such that if

$$m \geq C, \quad s \leq C' m^{1/(1+\eta)},$$

then the following holds with probability exceeding  $1 - \varepsilon - e^{-d^3 \log(3m)}$  with respect to the initial condition  $X^0 = x \sim d\mu$ : The polynomial coefficients of  $f$ , as well as the outlier vectors  $(\Theta, \Theta')$ , can be exactly recovered from the unique solution to the partial  $\ell_1$ -minimization problem

$$\min_{c, e} \|e\|_1 \text{ subject to } v - Ac = e.$$

We turn to the proof of Theorem 3.11 shortly. First, we provide a corollary of the theorem in case we observe measurements of the time- $\Delta$  map, for  $\Delta < 1$ .

**COROLLARY 3.12.** Fix  $\eta > 0, \varepsilon > 0, L \in \mathbb{N}$  such that  $\Delta = \frac{1}{L}$ , and degree  $p$ . Let  $X^t = x(t) = (x_1(t), x_2(t), x_3(t))$  be the flow generated by a vector field  $G \in \mathcal{U}$  whose governing equation  $f: \mathbb{R}^3 \rightarrow \mathbb{R}^3$  in  $\dot{x}(t) = f(x(t))$  is a multivariate algebraic polynomial of degree at most  $p$ , and suppose we observe corrupted measurements of the time- $\Delta$  map

$$U^j = X^{\Delta j} + \Theta_j, \quad V^j = \dot{X}^{\Delta j} + \Theta'_j, \quad j = 0, 1, 2, \dots, mL,$$

where  $(\Theta, \Theta') \in \mathbb{R}^{6m}$  is sparse and  $\|(\Theta, \Theta')\|_0 \leq 2sL$ .

Then for the same constants  $C, C'$  as in Theorem 3.11, once

$$m \geq C, \quad s \leq C' m^{1/(1+\eta)},$$

then the following holds with probability exceeding  $1 - \varepsilon - e^{-d^3 \log(3m)}$  with respect to  $x \sim d\mu$ : The polynomial coefficients of  $f$ , as well as the outlier vectors  $(\Theta, \Theta')$ , can be exactly recovered as the unique solution to the partial  $\ell_1$ -minimization problem

$$\min_{c, e} \|e\|_1 \text{ subject to } v - Ac = e.$$

*Proof of Corollary 3.12.* Consider the subsequences

$$U_k^j = U_k^{Lj+k}, V_k^j = V_k^{Lj+k}, \quad k = 0, 1, \dots, L-1, \quad j = 0, 1, 2, \dots, m.$$

Each of these  $L$  subsequences represents a corrupted measurement vector for a time-1 map of the flow  $X^t$ . By the pigeonhole principle, one of these subsequences is sparsely corrupted, having associated corrupted vector of sparsity level  $\|(\Theta_k, \Theta'_k)\|_0 \leq s$ . Thus we may apply Theorem 3.11, using Corollary 3.5 in place of Corollary 3.4.  $\square$

*Proof of Theorem 3.11.* We break the proof into several parts. First, consider a fixed polynomial coefficient vector  $c = (c_1, c_2, c_3) \in \mathbb{R}^{3r}$  of unit norm  $\|c\|_1 = 1$ , and the corresponding vector  $v^c(x) = (v_i^c(x_1, x_2, x_3))_{i=1}^3$  whose components are given by

$$v_i^c(x_1, x_2, x_3) = \sum_{\alpha: |\alpha| \leq p} c_i(\alpha) x_1^{\alpha_1} x_2^{\alpha_2} x_3^{\alpha_3}.$$

Consider the corresponding  $C^{1+\eta}$  observable  $\psi = \psi^c: \mathbb{R}^3 \rightarrow \mathbb{R}$  given by

$$(21) \quad \psi^c(x) = |v_1^c(x)|^{1+\eta} + |v_2^c(x)|^{1+\eta} + |v_3^c(x)|^{1+\eta}.$$

Combining the chain rule, Holder's inequality, and that  $||x|^\eta - |y|^\eta| \leq ||x| - |y||^\eta$  by concavity, we bound the  $C^{1+\eta}$  norm of  $\psi$  by

$$\begin{aligned} \|\psi^c\|_{C^{1+\eta}} &\leq \max_{|\beta| \leq 1} \sup_{x \in \Lambda} |D^\beta \psi^c(x)| + \max_{|\beta| \leq 1} \sup_{x \in \Lambda} |D^\beta \psi^c(x)| \\ &\leq 2\|c\|_1 (1+\eta)p ((B_\Lambda)^{2p-1})^\eta \\ &= 2(1+\eta)p ((B_\Lambda)^{2p-1})^\eta \\ &=: C_{\eta, \Lambda, p}. \end{aligned}$$

In particular,  $\|\psi^c\|_{C^{1+\eta}}$  is bounded above by a constant which is independent of the number of measurements  $m$ . Indeed, this bound holds uniformly over all observables corresponding to  $c: \|c\|_1 = 1$ :

$$(22) \quad \max_{c: \|c\|_1 = 1} \|\psi^c\|_{C^{1+\eta}} \leq C_{\eta, \Lambda, p}.$$

We also have a uniform lower bound on a related quantity by Lemma 3.10: Let  $D = D_{\eta, \Lambda, p} > 0$  be the lower bound in Lemma 3.10. Then

$$(23) \quad \min_{c: \|c\|_1=1} \int_{\Lambda} \psi^c(x) d\mu(x) \geq D_{\eta, \Lambda, p}.$$

Recall that  $m$  is our number of measurements. In order to apply a variant of Corollary 3.9 uniformly over all observables  $\{\psi^c : \|c\|_1 = 1\}$ , we first discretize the set  $\{c : \|c\|_1 = 1\}$  using covering lemmas, and then apply a large deviations result. By a well-known result in the literature on covering numbers (see, e.g., [20, Appendix C.2]), there exists a finite set of points  $\mathcal{Q}$  in  $\{c : \|c\|_1 = 1\}$  such that

$$\max_{c: \|c\|_1=1} \min_{q \in \mathcal{Q}} \|c - q\|_1 \leq 1/m,$$

of cardinality

$$|\mathcal{Q}| \leq (3m)^{3r}.$$

We now apply Corollary 3.9 of Proposition 3.7 uniformly over the observables  $\psi^q : q \in \mathcal{Q}$ . Draw an initial condition  $X^0 = x$  from the measure  $d\mu$ . The following holds with probability  $1 - \varepsilon$  with respect to the draw of  $x$ :

$$\begin{aligned} & \min_{q \in \mathcal{Q}} \left| \sum_{j=0}^{m-1} \psi^q(X^j) \right| \\ & \geq \min_{q \in \mathcal{Q}} \left| m \int_{\Lambda} \psi^q(x) d\mu(x) \right| - \max_{q \in \mathcal{Q}} \left| \sqrt{C_1} C_{\eta, \Lambda, p} \sqrt{m} Z^q \right| \\ & \quad - \max_{q \in \mathcal{Q}} \kappa_{\varepsilon, p, \eta} m^{1/4} (\log(m))^{1/2} (\log \log m)^{1/4} \\ & \geq m D_{\eta, \Lambda, p} - \sqrt{m} \sqrt{C_1} C_{\eta, \Lambda, p} \max_{q \in \mathcal{Q}} |Z^q| - \kappa_{\varepsilon, p, \eta} m^{1/4} (\log(m))^{1/2} (\log \log m)^{1/4}, \end{aligned}$$

where  $Z^q$  denotes a standard normal random variable. We now bound  $\max_{q \in \mathcal{Q}} |Z^q|$ . First, recall the Chernoff bound for a standard Gaussian random variable  $Z$ :

$$P(|Z| \geq t) \leq 2e^{-t^2/2} \quad \forall t \geq 0.$$

Recalling that  $|\mathcal{Q}| \leq (3n)^{3r}$ , the union bound then gives

$$P(\exists q \in \mathcal{Q} : |Z^q| \geq t) \leq 2(3m)^{3r} e^{-t^2/2} \leq 2e^{3r \log(3m) - t^2/2} \quad \forall t \geq 0.$$

In particular,

$$P(\forall q \in \mathcal{Q} : |Z^q| \leq 2\sqrt{3r \log(3m)}) \geq 1 - 2e^{-3r \log(3m)}.$$

All together, we find with probability exceeding  $1 - \varepsilon - 2e^{-3r \log(3m)}$  with respect to  $X^0 \sim d\mu$ ,

$$\min_{q \in \mathcal{Q}} \left| \sum_{j=0}^{m-1} \psi^q(X^j) \right| \geq mD - \sqrt{m \log(m)} C - \kappa m^{1/4} (\log(m))^{1/2} (\log \log m)^{1/4},$$

where the constants  $D, C$ , and  $\kappa$  depend only on  $\varepsilon, \eta, \Lambda$ , and  $p$ . Thus, for a sufficiently large constant  $C' = C'(\varepsilon, d, \eta)$  and sufficiently small constant  $C'' = C''(\varepsilon, d, \eta)$ , and

using the inequality  $\|v\|_1 \geq \|v\|_{1+\eta}$ , the following uniform lower bound holds with probability exceeding  $1 - \varepsilon - 2e^{-3r \log(3m)}$  with respect to  $X^0 \sim d\mu$ : If the number of measurements satisfies

$$(24) \quad m \geq C',$$

then

$$(25) \quad \min_{q \in \mathcal{Q}} \left( \sum_{j=0}^{m-1} |v_1^q(x)| + |v_2^q(x)| + |v_3^q(x)| \right) \geq (C''m)^{1/(1+\eta)}.$$

Recall now by (19) that

$$\sum_{j=1}^{3m} |(Aq)_j| = \sum_{j=0}^{m-1} \psi^q(X^j) + \sum_{j=1}^m \sum_{l=1}^3 |R_l(j)|.$$

Assuming (24) and with the same probability, and using the bound (20) on  $\|R_l\|_\infty$  and that  $\|R_l\|_0 \leq 2s$ ,

$$\begin{aligned} \min_{q \in \mathcal{Q}} \sum_{j=1}^{3m} |(Aq)_j| &\geq (C''m)^{1/(1+\eta)} - \sum_{j=1}^m \sum_{l=1}^3 |R_l(j)| \\ &\geq (C''m)^{1/(1+\eta)} - 6s(B_\Theta + B_\Lambda)^d. \end{aligned}$$

We now use a continuity argument to pass this lower bound from the discrete net  $\mathcal{Q}$  to the entire sphere  $\{c : \|c\|_1 = 1\}$ . Fix  $c \in \{c : \|c\|_1 = 1\}$  and let  $q \in \mathcal{Q}$  be such that  $\|c - q\|_1 \leq 1/m$ , which exists by construction of  $\mathcal{Q}$ . By Holder's inequality,

$$\begin{aligned} \|A(c - q)\|_1 &\leq 3mB_\Lambda^d \|c - q\|_1 + s(B_\Lambda + B_\Theta)^d \|c - q\|_1 \\ (26) \quad &\leq 3B_\Lambda^d + 3\frac{s}{m}(B_\Lambda + B_\Theta)^d. \end{aligned}$$

Thus,

$$\min_{c: \|c\|_1=1} \|Ac\|_1 \geq (C''m)^{1/(1+\eta)} - C'''s(B_\Lambda + B_\Theta)^d.$$

At the same time, for a subset  $S \subset \{1, 2, \dots, m\}$  of size  $s$ , we have the immediate and uniform upper bound

$$\sum_{j \in S} |(Ac)_j| \leq sB_\Lambda^d \|c\|_1.$$

Thus, there is a constant  $C''''$  depending on only  $\varepsilon, p, \eta$ , and  $B_\Theta$  such that if  $s \leq C''''m^{1/(1+\eta)}$ , then uniformly over  $c : \|c\|_1 = 1$ , and uniformly over subsets  $S \subset [3m]$  of size  $s$ ,

$$(27) \quad \|(Ac)_S\|_1 < \frac{1}{2} \|Ac\|_1,$$

and in particular,  $\|Ac\|_1 > 0$  for all  $c : \|c\|_1 = 1$ , implying that the columns of  $A$  are linearly independent.

The theorem follows for  $c \neq 0$  of arbitrary  $\ell_1$  norm by normalizing both sides of the expression (27). The theorem follows by application of Corollary 3.4.  $\square$

As a consequence of our proof, we also provide theoretical guarantees for the algorithm of [7] in the noiseless case.

**COROLLARY 3.13.** *Under the same conditions as in Theorem 3.11, the matrix  $\Phi$  constructed from uncorrupted measurements of the time- $\Delta$  map  $X_j = X^{\Delta_j}$  is full rank, provided that  $m \geq C'$ .*

**4. Numerical method.** In this section, we explain how to solve our proposed model (9) numerically by using augmented Lagrangian/Bregman distance and alternating minimization method. Recall our proposed model:

$$\begin{aligned} \min_{(\mathcal{C}, \mathcal{E})} \|\mathcal{E}\|_{2,1} &= \min_{(\mathcal{C}, \mathcal{E})} \sum_{j=1}^m \|\mathcal{E}(j, :)\|_2, \\ \text{s.t. } \Phi(X^\circ)\mathcal{C} + \mathcal{E} &= \dot{X}^\circ \quad \text{and } \mathcal{C} \text{ is sparse.} \end{aligned}$$

The corresponding augmented Lagrangian is of the form

$$\begin{aligned} (28) \quad (\mathcal{C}^{k+1}, \mathcal{E}^{k+1}) &= \min_{(\mathcal{C}, \mathcal{E})} \sum_{j=1}^m \|\mathcal{E}(j, :)\|_2 + \frac{\mu}{2} \|\Phi(X^\circ)\mathcal{C} + \mathcal{E} - \dot{X}^\circ + b^k\|_F^2, \\ b^{k+1} &= b^k + \Phi(X^\circ)\mathcal{C}^{k+1} + \mathcal{E}^{k+1} - \dot{X}^\circ. \end{aligned}$$

Now we can apply the alternating minimization method to solve problem (28).

- The  $\mathcal{C}$ -subproblem:

$$\mathcal{C}^{k+1} = \min_{\mathcal{C}} \|\Phi(X^\circ)\mathcal{C} + \mathcal{E}^k - \dot{X}^\circ + b^k\|_F^2, \quad \text{s.t. } \mathcal{C} \text{ is sparse.}$$

- The  $\mathcal{E}$ -subproblem:

$$\mathcal{E}^{k+1} = \min_{\mathcal{E}} \sum_{j=1}^m \|\mathcal{E}(j, :)\|_2 + \frac{\mu}{2} \|\Phi(X^\circ)\mathcal{C}^{k+1} + \mathcal{E} - \dot{X}^\circ + b^k\|_F^2.$$

Notice that the  $\mathcal{C}$ -subproblem is a least-squares problem for an overdetermined system. The sparsity of  $\mathcal{C}$  is a property of the system itself; therefore, we enforce the sparsity of  $\mathcal{C}$  by applying the hard-thresholding operator after obtaining the least-squares solution:

$$\mathcal{C}^{k+1} = S_h \left( (\Phi(X^\circ))^{-1} (\dot{X}^\circ - \mathcal{E}^k - b^k), \lambda \right),$$

where

$$S_h(u, \gamma) := u \cdot I_{|u| \geq \gamma} = \begin{cases} u & \text{if } |u| \geq \gamma, \\ 0 & \text{otherwise.} \end{cases}$$

It is also discussed in [7] that the mentioned method is robust to noise in recovering the true coefficients  $\mathcal{C}$ .

The solution for  $\mathcal{E}$  is given explicitly:

$$\mathcal{E}^{k+1} = S_2 \left( \dot{X}^\circ - b^k - \Phi(X^\circ)\mathcal{C}^{k+1}, \mu \right),$$

where

$$S_2(u_j, \gamma) = \max \left( 1 - \frac{1}{\gamma \|u_j\|_2}, 0 \right) u_j,$$

for each row  $u_j$  of  $u$ .

Below is the summary of the algorithm for problem (28).

**Algorithm**

Given:  $\mathcal{E}^0, b^0, tol$ , and parameters  $\lambda, \mu$ .

**while**  $\|\mathcal{E}^k - \mathcal{E}^{k-1}\|_\infty > tol$  **do**

$$\mathcal{C}^{k+1} = S_h \left( (\Phi(X^\circ))^{-1}(\dot{X}^\circ - \mathcal{E}^k - b^k), \lambda \right)$$

$$\mathcal{E}^{k+1} = S_2 \left( \dot{X}^\circ - b^k - \Phi(X^\circ)\mathcal{C}^{k+1}, \mu \right)$$

$$b^{k+1} = b^k + \Phi(X^\circ)\mathcal{C}^{k+1} + \mathcal{E}^{k+1} - \dot{X}^\circ$$

**end while**

**5. Numerical results.** In this section, we apply the method from section 4 to various chaotic systems including the well-known Lorenz system and Rössler system. Moreover, our proposed reconstruction method also works numerically for systems exhibiting so-called *hyperchaos* whose dimension is greater than three, which suggests that our reconstruction guarantee may extend to higher-dimensional systems. We reiterate that chaotic systems are not only well-suited for theoretical reconstruction guarantees, but also regimes where it is of utmost importance to recover the governing equations with high precision, in light of the property of sensitivity to initial conditions. Therefore, we define the following relative error formula for the coefficients:

$$\text{coefficient error} := \max \left( \max_{\mathcal{C}_{\text{true}}(i) \neq 0} \left| \frac{\mathcal{C}_{\text{recovered}}(i) - \mathcal{C}_{\text{true}}(i)}{\mathcal{C}_{\text{true}}(i)} \right|, \max_{\substack{\mathcal{C}_{\text{true}}(i)=0, \\ \mathcal{C}_{\text{recovered}}(i) \neq 0}} |\mathcal{C}_{\text{recovered}}(i)| \right).$$

In other words, the coefficient error measures the maximal relative recovery error over the different polynomial coefficients. Throughout this section, error stands for coefficient error, unless otherwise stated.

In practice, we will not observe the derivative information  $\dot{X}^\circ$ . Instead, we can approximate the rate of change in the system from the state space measurements using first-order, second-order, or higher-order approximations. Explicitly, given  $x(t) \in \mathbb{R}^n$ , the rate of change in  $x$  can be approximated as

$$\begin{aligned} \dot{x}(t) &= \frac{x(t+dt) - x(t)}{dt} + \mathcal{O}(dt) \quad (\text{first-order approximation}), \\ \dot{x}(t) &= \frac{x(t+dt) - x(t-dt)}{2dt} + \mathcal{O}(dt^2) \quad (\text{second-order approximation}). \end{aligned}$$

To get better accuracy, we use the second-order approximation of rate of change for our optimization model.

We first show some numerical results for the Lorenz system

$$(29) \quad \begin{cases} \frac{dx_1}{dt} &= 10(x_2 - x_1), \\ \frac{dx_2}{dt} &= x_1(28 - x_3) - x_2, \\ \frac{dx_3}{dt} &= x_1x_2 - \frac{8}{3}x_3, \end{cases}$$

with different percentages of corruption. To simulate the measurement data, we first solve the Lorenz system (29) using the fourth-order Runge–Kutta method with  $dt = 0.0005$ . Then we randomly assign locations where the data is corrupted over intervals of time, and add Gaussian noise with standard deviation  $\sigma$  to the data at those corrupted intervals. The bandwidth of corruption ranges from 5 to 50. From the simulated data, we build the matrix  $X^\circ$ , compute the time derivative  $\dot{X}^\circ$  using the second-order approximation of derivatives, and build the dictionary  $\Phi(X^\circ)$ . We verify our algorithm for the integration (from  $t = 0$  to  $t = 20$ ,  $dt = 0.0005$ ) of 40000 measurements with different percentages of corruption. The results are shown in Figures 1 and 2. In all cases, our algorithm can detect the exact locations of the outliers and recover the coefficients in the polynomial equations with very high accuracy. Notice that our model can tolerate a high percentage of corruption as long as the size of the data is sufficiently large.

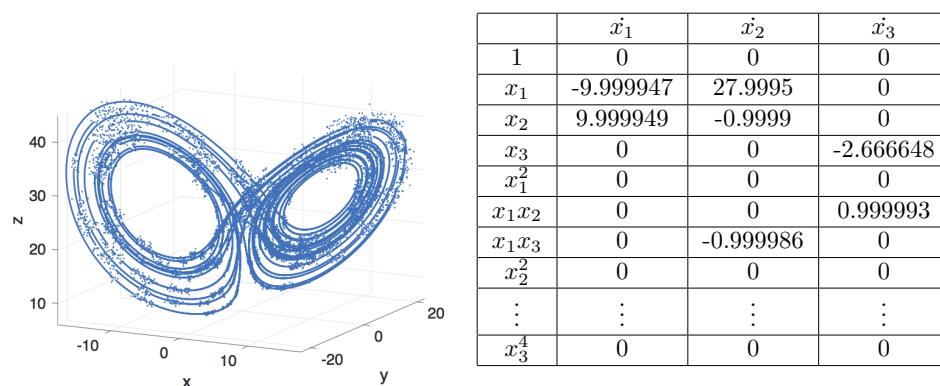


FIG. 1. Left: Lorenz system (29), with 19.19% corrupted data,  $T_{\text{final}} = 20$ ,  $dt = 0.0005$ ,  $\text{hard-thres} = 0.1$ ,  $\text{row-thres} = 0.0125$ ,  $\text{tol} = 0.005$ . Right: The recovered coefficients. The model recovers the coefficients within 0.0096% error and detects the exact locations of the outliers after 22 iterations.

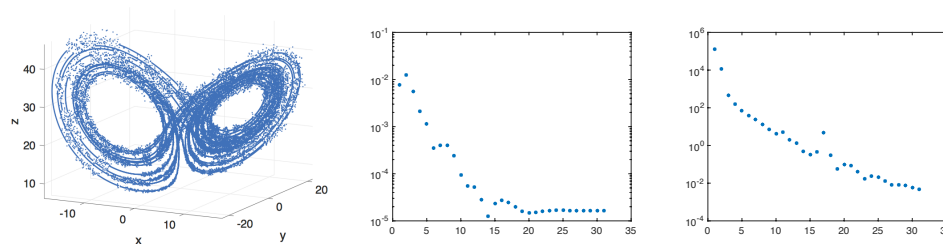


FIG. 2. Left: Lorenz system (29), with 49.53% corrupted data,  $T_{\text{final}} = 20$ ,  $dt = 0.0005$ ,  $\text{hard-thres} = 0.1$ ,  $\text{row-thres} = 0.0125$ ,  $\text{tol} = 0.005$ . The error between the recovered coefficients iterates and the true ones (middle) and the error between two consecutive  $\mathcal{E}$  (right) versus the number of iterations (in logarithmic scale for the vertical axis). The model recovers the coefficients with 0.0096% error and detects the exact locations of the outliers after 31 iterations.

We also examine our proposed scheme for the Lorenz system over a shorter interval of time. In Figure 3, 5000 measurements are given with 22.55% and 71.89% corrupted, corresponding to  $dt = 0.0005$  and  $T_{\text{final}} = 2.5$ . All other parameters remain the same as in the previous example. The model recovers the coefficients within 0.0317% and 0.0477% errors, respectively, and detects the exact locations of outliers after 24



iterations. If the number of measurements is smaller, for example, if  $T_{\text{final}} \leq 2$ , then the scheme does not work as well, illustrating the importance of observing the system over a critical minimal amount of time.

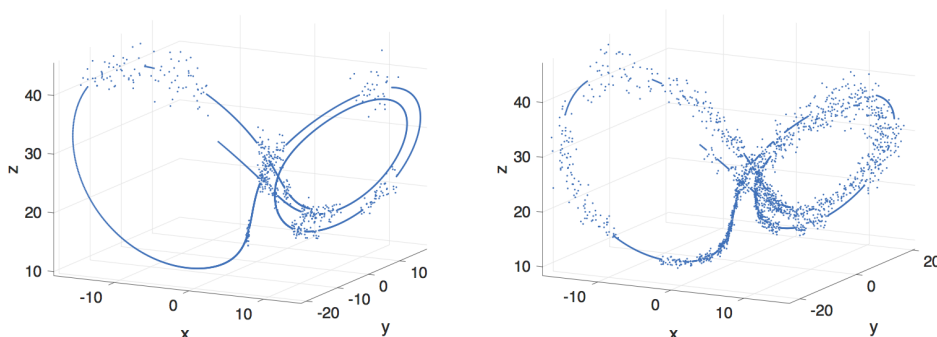


FIG. 3. *Lorenz trajectories with 22.55% corruption (left) and with 71.89% corruption (right), associated with the system (29),  $T_{\text{final}} = 2.5$ ,  $dt = 0.001$ ,  $\text{hard-thres} = 0.1$ ,  $\text{row-thres} = 0.0125$ ,  $\text{tol} = 0.005$ . The model detects the exact locations of the outliers in both data and recovers the coefficients with the errors 0.0317% and 0.0477%, respectively.*

Finally, we verify our proposed scheme for the Lorenz system over a much longer interval of time, in order to gauge whether the butterfly effect (sensitivity to initial conditions) begins to interfere with the recovery conditions. In Figure 4, 200000 measurements are taken with 19.75% corruption, corresponding to  $dt = .0005$  and  $T_{\text{final}} = 100$  and the other parameters as in previous examples. The model recovers the coefficients within 0.0097% error and detects the exact locations of the outliers after 29 iterations. In the next step, we show the robustness of our model to a small amount of additional additive noise. After simulating the corrupted data from the Lorenz system, we add Gaussian noise to the entire data. From the noisy data, we build the dictionary and approximate the time derivative. To test the robustness of our model, we simulate 100 sets of data and check how many times our model can detect the exact locations of outliers. The summary is shown in Table 1.

TABLE 1

*Different noise levels and the recovery results associated with the Lorenz system,  $T_{\text{final}} = 20$ ,  $dt = 0.0005$ ,  $\text{hard-thres} = 0.1$ ,  $\text{row-thres} = 0.0125$ ,  $\text{tol} = 0.005$ , and around 20% corrupted.*

$\sigma_{\text{noise}}$	# Times (out of 100) outliers detected exactly	Coefficient error (%)
$0.4*dt$	89	min = 0.0009, max = 0.0525
$0.6*dt$	87	min = 0.0006, max = 0.9395
$0.8*dt$	65	min = 0.0012, max = 1.57

Next, we examine the Rössler system

$$(30) \quad \begin{cases} \frac{dx_1}{dt} = -x_2 - x_3, \\ \frac{dx_2}{dt} = x_1 + 0.2x_2, \\ \frac{dx_3}{dt} = 0.2 - 5.7x_3 + x_1x_3, \end{cases}$$

with different percentages of corruption. We simulate the data in the same way as for the Lorenz system. Our algorithm can detect exactly the locations of outliers in all

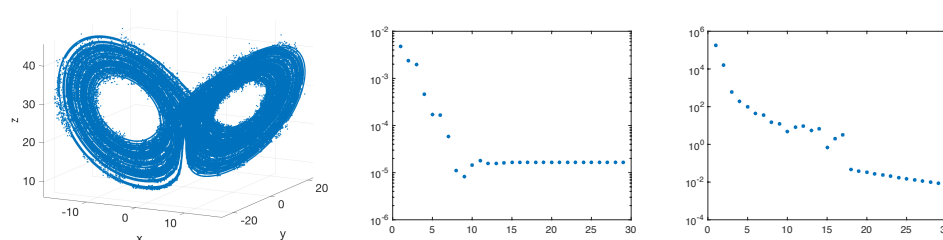
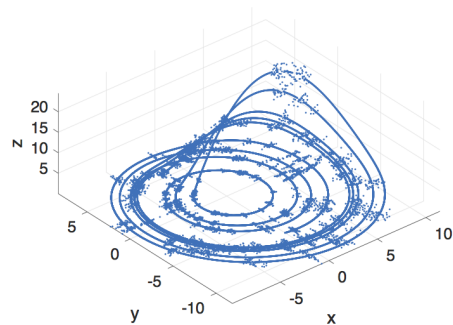


FIG. 4. Left: Lorenz system (29), with 19.75% corrupted data,  $T_{\text{final}} = 100$ ,  $dt = 0.0005$ ,  $\text{hard-thres} = 0.1$ ,  $\text{row-thres} = 0.0125$ ,  $\text{tol} = 0.01$ . The error between the recovered coefficients iterates and the true ones (middle) and the error between two consecutive  $\mathcal{E}$  (right) versus the number of iterations (in logarithmic scale for the vertical axis). The model recovers the coefficients with 0.0097% error and detects the exact locations of the outliers after 29 iterations.



	$\dot{x}_1$	$\dot{x}_2$	$\dot{x}_3$
1	0	0	0.2012
$x_1$	0	0.9994	0
$x_2$	-1.0021	0.1948	0
$x_3$	-1.0044	0	-5.7009
$x_1^2$	0	0	0
$x_1x_2$	0	0	0.00
$x_1x_3$	0	0	1.0009
$x_2^2$	0	0	0
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$x_3^4$	0	0	0

FIG. 5. Left: Rössler system (30) with 10.2% corrupted data,  $T_{\text{final}} = 50$ ,  $dt = 0.0005$ ,  $\text{hard-thres} = 0.05$ ,  $\text{row-thres} = 0.025$ ,  $\text{tol} = 10^{-5}$ . Right: The recovered coefficients. The model recovers the coefficients with 0.6% error and detect the exact locations of the outliers after 160 iterations.

such cases, and recover the constant coefficient as well as other coefficients with very high accuracy. The numerical results are shown in Figures 5, 6, and 7.

In our final example, we apply our model to the following hyperchaos [39]

$$(31) \quad \begin{cases} \frac{dx_1}{dt} = -x_2 - x_3, \\ \frac{dx_2}{dt} = x_1 + 0.25x_2 + x_4, \\ \frac{dx_3}{dt} = 3 + x_1x_3, \\ \frac{dx_4}{dt} = -0.5x_3 + 0.05x_4. \end{cases}$$

As mentioned in [39], the variable  $x_3(t)$  serves to check the growth of the flow  $(x_1, x_2, x_4)$  from time to time. We numerically solve the system (31) using the fourth Runge–Kutta method with  $dt = 0.001$ ,  $T_{\text{final}} = 100$ . The flow of the 3D subspace  $(x_1, x_2, x_4)$  is presented in Figure 8 along with the remaining variable  $x_3$  plot. Now we randomly assign the locations where the data is corrupted, and add Gaussian noise to the data at those corrupted intervals; see Figure 9. The numerical result is shown in Table 2.

**6. Conclusion and discussion.** Using statistical properties of Lorenz-like systems and partial sparse recovery guarantees from compressive sensing, we provide

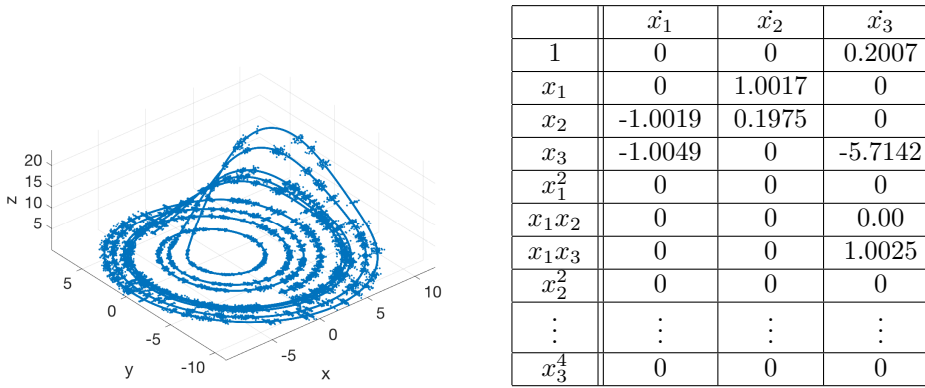


FIG. 6. Left: Rössler system (30) with 20% corrupted data,  $T_{final} = 50$ ,  $dt = 0.0005$ ,  $hard-thres = 0.05$ ,  $row-thres = 0.025$ . Right: The recovered coefficients. The model recovers the coefficients with 1.25% error and detect the exact locations of the outliers after 160 iterations.

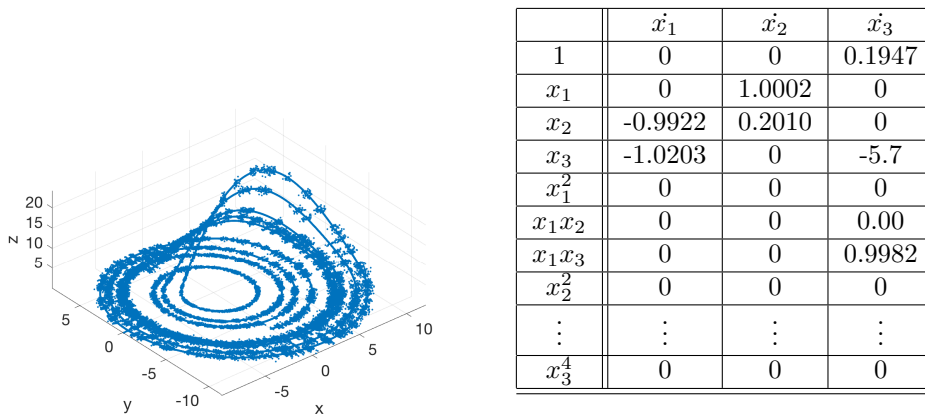


FIG. 7. Left: Rössler system (30) with 40% corrupted data,  $T_{final} = 50$ ,  $dt = 0.0005$ ,  $hard-thres = 0.05$ ,  $row-thres = 0.025$ . Right: The recovered coefficients. The model recovers the coefficients with 2% error and detects the exact locations of the outliers after 314 iterations.

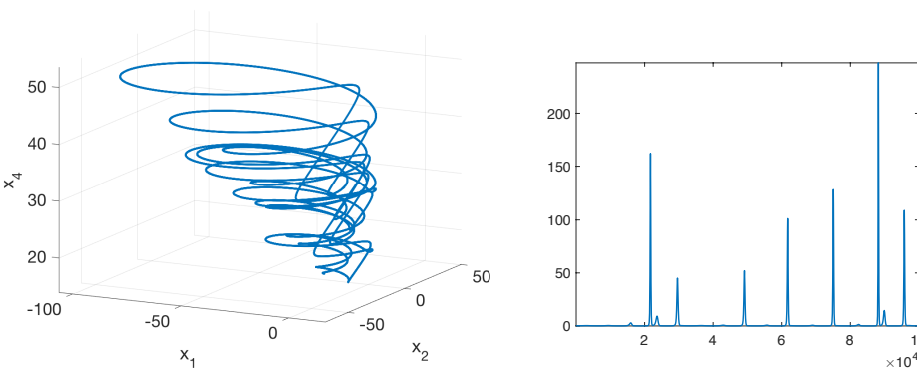


FIG. 8. Left: Plot of trajectory flow in  $(x_1, x_2, x_4)$  subspace. Right: Plot of  $x_3$  along time associated with the hyperchaos (31).  $T_{final} = 100$ ,  $dt = 0.001$ .

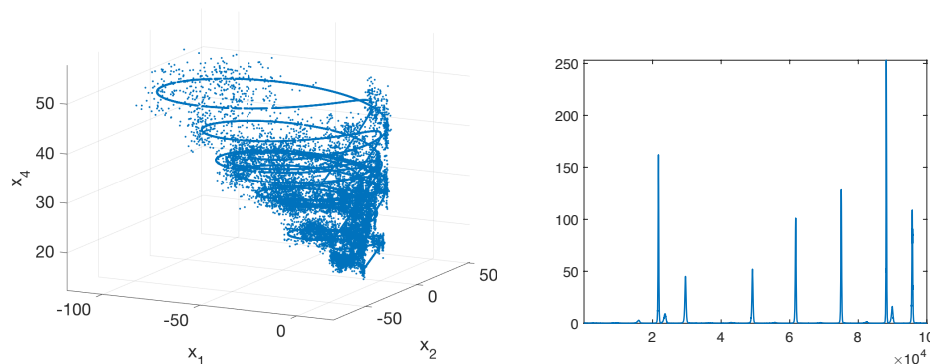


FIG. 9. Rössler hyperchaos system (30) with 10.25% corrupted data,  $T_{\text{final}} = 100$ ,  $dt = 0.001$ .

TABLE 2

The recovered coefficients table for data associated with Figure 9. The model recovers the coefficients and detect exactly the locations of the outliers after 201 iterations.

	$\dot{x}_1$	$\dot{x}_2$	$\dot{x}_3$	$\dot{x}_4$
1	0	0	2.9998	0
$x_1$	0	0.99999	0	0
$x_2$	-0.99999	0.24999	0	0
$x_3$	-0.99999	0	0	-0.49999
$x_4$	0	0.99999	0	0.04999
$x_1^2$	0	0	0	0
$x_1 x_2$	0	0	0	0
$x_1 x_3$	0	0	0.99999	0
$x_1 x_4$	0	0	0	0
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$x_4^4$	0	0	0	0

conditions for recovering the governing equations from possibly highly corrupted measurement data. In addition, a stable numerical scheme is presented to recover the coefficients of the underlying equations in the space of multivariable polynomials with high accuracy and exactly identify the outliers, despite being in a regime of sensitivity to initial conditions. Our method might be useful for recovering the governing equations more generally, when the governing equations do not necessarily have a polynomial form. Explicitly, by doing our method “locally,” we can approximate the Taylor series expansion to the governing equations (assuming they are sufficiently smooth), and then “piece together” the recovered polynomial Taylor expansions to approximate smooth governing equations more generally. In the future, we also would like to adjust the numerical scheme and theoretical guarantees to handle data with much higher levels of noise in addition to outliers. We are also interested in extending theory and algorithm to settings where we observe snapshots of an observable of phase space (rather than the phase space measurement in its entirety) in the presence of outliers, as well as explore the sparsity structures in other high-dimensional nonlinear functional spaces such as the space of Legendre polynomials [38] and higher-dimensional spaces. Finally, the constants in our theorem such that  $m \geq C$  and  $s \leq C' m^{1/(1+\eta)}$  are not explicit. This follows because the constants are not explicit in the current theory for dynamical systems. So perhaps our work can motivate researchers in dynamical

systems to make the associated constants more explicit.

**Acknowledgments.** We would like to thank Afonso Bandeira and Stefan Steinerberger for helpful comments that improved this paper.

#### REFERENCES

- [1] L. A. AGUIRRE AND C. LETELLIER, *Modeling nonlinear dynamics and chaos: a review*, Math. Probl. Eng., 2009 (2009), 238960.
- [2] V. ARAUJO, S. GALATOLO, AND M. J. PACIFICO, *Statistical properties of Lorenz-like flows, recent developments and perspectives*, Internat. J. Bifur. Chaos Appl. Sci. Engrg., 24 (2014), 1430028.
- [3] V. ARAÚJO AND I. MELBOURNE, *Exponential decay of correlations for nonuniformly hyperbolic flows with a  $C^{1+\alpha}$  stable foliation, including the classical Lorenz attractor*, Ann. Henri Poincaré, 17 (2016), pp. 2975–3004.
- [4] V. ARAUJO, I. MELBOURNE, AND P. VARANDAS, *Rapid mixing for the Lorenz attractor and statistical limit laws for their time-1 maps*, Comm. Math. Phys., 340 (2015), pp. 901–938.
- [5] A. S. BANDEIRA, K. SCHEINBERG, AND L. N. VICENTE, *On partial sparse recovery*, preprint 11-13, Department of Mathematics, University of Coimbra, Coimbra, Portugal, 2011.
- [6] I. BRIGHT, G. LIN, AND J. N. KUTZ, *Compressive sensing based machine learning strategy for characterizing the flow around a cylinder with limited pressure measurements*, Phys. Fluids, 25 (2013), 127102.
- [7] S. L. BRUNTON, J. L. PROCTOR, AND J. N. KUTZ, *Discovering governing equations from data by sparse identification of nonlinear dynamical systems*, Proc. Natl. Acad. Sci. USA, 113 (2016), pp. 3932–3937.
- [8] S. L. BRUNTON, J. H. TU, I. BRIGHT, AND J. N. KUTZ, *Compressive sensing and low-rank libraries for classification of bifurcation regimes in nonlinear dynamical systems*, SIAM J. Appl. Dyn. Syst., 13 (2014), pp. 1716–1732, <https://doi.org/10.1137/130949282>.
- [9] R. E. CAFLISCH, S. J. OSHER, H. SCHAEFFER, AND G. TRAN, *PDEs with compressed solutions*, Commun. Math. Sci., 13 (2015), pp. 2155–2176.
- [10] E. J. CANDLES, J. K. ROMBERG, AND T. TAO, *Stable signal recovery from incomplete and inaccurate measurements*, Comm. Pure Appl. Math., 59 (2006), pp. 1207–1223.
- [11] M. CASDAGLI, S. EUBANK, J. D. FARMER, AND J. GIBSON, *State space reconstruction in the presence of noise*, Phys. D, 51 (1991), pp. 52–98.
- [12] M. CHENG, T. Y. HOU, AND Z. ZHANG, *A dynamically bi-orthogonal method for time-dependent stochastic partial differential equations I: Derivation and algorithms*, J. Comput. Phys., 242 (2013), pp. 843–868.
- [13] A. COHEN, W. DAHMEN, AND R. DEVORE, *Compressed sensing and best  $k$ -term approximation*, J. Amer. Math. Soc., 22 (2009), pp. 211–231.
- [14] J. P. CRUTCHFIELD AND B. S. MCNAMARA, *Equations of motion from a data series*, Complex Systems, 1 (1987), pp. 417–452.
- [15] T. S. CUBITT, J. EISERT, AND M. M. WOLF, *Extracting dynamical equations from experimental data is NP hard*, Phys. Rev. Lett., 108 (2012), 120503.
- [16] W. DENG, W. YIN, AND Y. ZHANG, *Group sparse optimization by alternating direction method*, in Proceedings of SPIE Optical Engineering+ Applications, International Society for Optics and Photonics, San Diego, CA, 2013, 88580R.
- [17] D. DONOHO, *Compressed sensing*, IEEE Trans. Inform. Theory, 52 (2006), pp. 1289–1306.
- [18] M. FORNASIER AND H. RAUHUT, *Recovery algorithms for vector-valued data with joint sparsity constraints*, SIAM J. Numer. Anal., 46 (2008), pp. 577–613, <https://doi.org/10.1137/0606668909>.
- [19] M. FORNASIER, K. SCHNASS, AND J. VYBIRAL, *Learning functions of few arbitrary linear parameters in high dimensions*, Found. Comput. Math., 12 (2012), pp. 229–262.
- [20] S. FOUCART AND H. RAUHUT, *A Mathematical Introduction to Compressive Sensing*, Springer, New York, 2013.
- [21] T. HASTIE, R. TIBSHIRANI, AND M. WAINWRIGHT, *Statistical Learning with Sparsity: The Lasso and Generalizations*, Chapman and Hall/CRC Press, Boca Raton, FL, 2015.
- [22] M. JORDAN AND T. MITCHELL, *Machine learning: Trends, perspectives, and prospects*, Science, 349 (2015), pp. 255–260.
- [23] M. B. KENNEL, R. BROWN, AND H. D. ABARBANEL, *Determining embedding dimension for phase-space reconstruction using a geometrical construction*, Phys. Rev. A, 45 (1992), pp. 3403–3411.

- [24] M. KOWALSKI, *Sparse regression using mixed norms*, Appl. Comput. Harmon. Anal., 27 (2009), pp. 303–324.
- [25] D. KUGIUMTZIS, *State space reconstruction parameters in the analysis of chaotic time series – the role of the time window length*, Phys. D, 95 (1996), pp. 13–28.
- [26] C. LAINSCSEK, *A class of lorenz-like systems*, Chaos, 22 (2012), 013126.
- [27] L. LJUNG, *System Identification: Theory for the User*, Prentice Hall, Upper Saddle River, NJ, 1998.
- [28] L. LJUNG, *Perspectives on system identification*, Annu. Rev. Control, 34 (2010), pp. 1–12.
- [29] E. N. LORENZ, *Deterministic nonperiodic flow*, J. Atmos. Sci., 20 (1963), pp. 130–148.
- [30] J. LÜ, G. CHEN, AND D. CHENG, *A new chaotic system and beyond: the generalized Lorenz-like system*, Internat. J. Bifur. Chaos Appl. Sci. Engrg., 14 (2004), pp. 1507–1537.
- [31] A. MACKEY, H. SCHAEFFER, AND S. OSHER, *On the compressive spectral method*, Multiscale Model. Simul., 12 (2014), pp. 1800–1827, <https://doi.org/10.1137/140965909>.
- [32] V. OZOLIŅŠ, R. LAI, R. CAFLISCH, AND S. OSHER, *Compressed modes for variational problems in mathematics and physics*, Proc. Natl. Acad. Sci. USA, 110 (2013), pp. 18368–18373.
- [33] V. OZOLIŅŠ, R. LAI, R. CAFLISCH, AND S. OSHER, *Compressed plane waves yield a compactly supported multiresolution basis for the Laplace operator*, Proc. Natl. Acad. Sci. USA, 111 (2014), pp. 1691–1696.
- [34] N. H. PACKARD, J. P. CRUTCHFIELD, J. D. FARMER, AND R. S. SHAW, *Geometry from a time series*, Phys. Rev. Lett., 45 (1980), pp. 712–716.
- [35] J. PEREIRA, M. IBRAHIMI, AND A. MONTANARI, *Learning networks of stochastic differential equations*, in Proceedings of the 23rd International Conference on Neural Information Processing Systems, Vancouver, BC, 2010, pp. 172–180.
- [36] J. B. A. PERIERA AND M. IBRAHIMI, *Support recovery for the drift coefficient of high-dimensional diffusions*, IEEE Trans. Inform. Theory, 60 (2014), pp. 4026–4049.
- [37] J. PROCTOR, S. BRUNTON, B. BRUNTON, AND J. KUTZ, *Exploiting sparsity and equation-free architectures in complex systems*, Eur. Phys. J. Spec. Top., 223 (2014), pp. 2665–2684.
- [38] H. RAUHUT AND R. WARD, *Sparse Legendre expansions via  $\ell_1$ -minimization*, J. Approx. Theory, 164 (2012), pp. 517–533.
- [39] O. ROSSLER, *An equation for hyperchaos*, Phys. Lett. A, 71 (1979), pp. 155–157.
- [40] O. E. RÖSSLER, *Chemical turbulence: Chaos in a simple reaction-diffusion system*, Z. Naturforsch., 31A (1976), pp. 1168–1172.
- [41] S. T. ROWEIS AND L. K. SAUL, *Nonlinear dimensionality reduction by locally linear embedding*, Science, 290 (2000), pp. 2323–2326.
- [42] G. ROWLANDS AND J. SPROTT, *Extraction of dynamical equations from chaotic data*, Phys. D, 58 (1992), pp. 251–259.
- [43] H. SCHAEFFER, R. CAFLISCH, C. D. HAUCK, AND S. OSHER, *Sparse dynamics for partial differential equations*, Proc. Natl. Acad. Sci. USA, 110 (2013), pp. 6634–6639.
- [44] K. SCHITTKOWSKI, *Numerical Data Fitting in Dynamical Systems: A Practical Introduction with Applications and Software*, Appl. Optim. 77, Springer Science & Business Media, Dordrecht, 2013.
- [45] M. SCHMIDT AND H. LIPSON, *Distilling free-form natural laws from experimental data*, Science, 324 (2009), pp. 81–85.
- [46] J. SJÖBERG, Q. ZHANG, L. LJUNG, A. BENVENISTE, B. DELYON, P.-Y. GLORENNEC, H. HJALMARSSON, AND A. JUDITSKY, *Nonlinear black-box modeling in system identification: a unified overview*, Automatica J. IFAC, 31 (1995), pp. 1691–1724.
- [47] J. C. SPROTT, *Some simple chaotic flows*, Phys. Rev. E (3), 50 (1994), pp. R647–R650.
- [48] J. C. SPROTT AND S. J. LINZ, *Algebraically simple chaotic flows*, Int. J. Chaos Theor. Appl., 5 (2000), pp. 1–20.
- [49] F. TAKENS, *Detecting strange attractors in turbulence*, in Dynamical Systems and Turbulence, Warwick 1980, Lecture Notes in Math. 898, Springer, Berlin, New York, 1981, pp. 366–381.
- [50] G. TRAN, H. SCHAEFFER, W. M. FELDMAN, AND S. J. OSHER, *An  $l^1$  penalty method for general obstacle problems*, SIAM J. Appl. Math., 75 (2015), pp. 1424–1444, <https://doi.org/10.1137/140963303>.
- [51] W. TUCKER, *The Lorenz attractor exists*, C. R. Acad. Sci. Paris Sér. I Math., 328 (1999), pp. 1197–1202.
- [52] V. VAPNIK, *The Nature of Statistical Learning Theory*, Science & Business Media, New York, 2013.
- [53] D. VISWANATH, *The fractal property of the Lorenz attractor*, Phys. D, 190 (2004), pp. 115–128.
- [54] M. O. WILLIAMS, E. SHLIZERMAN, J. WILKENING, AND J. N. KUTZ, *The low dimensionality of time-periodic standing waves in water of finite and infinite depth*, SIAM J. Appl. Dyn. Syst., 11 (2012), pp. 1033–1061, <https://doi.org/10.1137/11084621X>.