RESEARCH ARTICLE | JUNE 18 2025

# Computing chaotic time-averages from few periodic or non-periodic orbits

Joshua L. Pughe-Sanford ; Sam Quinn ; Teodor Balabanski ; Roman O. Grigoriev



*Chaos* 35, 063133 (2025)
https://doi.org/10.1063/5.0264212
CHORUS

# Computing chaotic time-averages from few periodic or non-periodic orbits

Cite as: Chaos 35, 063133 (2025); doi: 10.1063/5.0264212
Submitted: 9 February 2025 · Accepted: 3 June 2025 ·
Published Online: 18 June 2025



Joshua L. Pughe-Sanford,[a] Sam Quinn, Teodor Balabanski, and Roman O. Grigoriev

## AFFILIATIONS

School of Physics, Georgia Institute of Technology, 837 State St NW, Atlanta, Georgia 30332, USA

[a]**Author to whom correspondence should be addressed:** jpughesanford@gatech.edu

## ABSTRACT

For appropriately chosen weights, temporal averages in chaotic systems can be approximated as a weighted sum of averages over reference states, such as unstable periodic orbits. Under strict assumptions, such as completeness of the orbit library, these weights can be formally derived using periodic orbit theory. When these assumptions are violated, weights can be obtained empirically using a Markov partition of the chaotic set. Here, we describe an alternative, data-driven approach to computing weights that allows for an accurate approximation of temporal averages from a variety of reference states, including both periodic orbits and non-periodic trajectory segments embedded within the chaotic set. For a broad class of observables, we demonstrate that the resulting reduced-order statistical description significantly outperforms those based on periodic orbit theory or Markov models, achieving superior accuracy while requiring far fewer reference states—two critical properties for applications to high-dimensional chaotic systems.





**Chaotic systems, which arise in a myriad of natural and engineered settings, challenge traditional methods of predicting long-term behavior due to their sensitivity to initial conditions. While periodic orbit theory provides a foundational framework for approximating time-averaged observables in chaotic systems, its reliance on complete sets of unstable periodic orbits (UPOs) often limits its applicability in high-dimensional settings where such complete sets may be very expensive or even impossible to identify. This work introduces a data-driven, interpretable approach leveraging reproducing kernel Hilbert space (RKHS) interpolation, termed least squares weighting (LSW), to estimate chaotic averages using arbitrary solution segments, including non-periodic "snippets." The LSW method achieves superior accuracy and faster convergence compared to state-of-the-art approaches such as periodic orbit theory, even with a limited library of states. By expanding the toolkit for analyzing high-dimensional chaos, our results open pathways for applications in fluid dynamics, magnetohydrodynamics, optics, and beyond, offering robust predictions in systems where traditional methods falter.**

## I. INTRODUCTION

Deterministic chaos is pervasive in natural and engineered systems, with turbulence in fluids,[1] plasma,[2] and nonlinear optics[3] representing some well-known and practically important examples of high-dimensional chaotic systems. Despite being governed by deterministic equations, chaotic systems can only be described dynamically on short temporal intervals due to their inherent sensitivity to initial conditions.[4] Hence, on long time scales, one is forced to resort to a statistical description, with a key objective of computing temporal averages of various physical observables. An illustrative example of dynamical vs statistical descriptions of chaotic systems is the prediction of weather vs the prediction of climate.

Here, we aim to predict the averages of physical observables, $a(\mathbf{x})$, defined as functions that map the instantaneous state of the system, $\mathbf{x} \in \mathscr{M}$, to a real number. So long as the dynamics are ergodic—a property many chaotic systems either have or may be presumed to have in practice—Birkhoff's ergodic theorem[5] ensures that temporal averages may be computed as the expectation value over a Sinai–Ruelle–Bowen (SRB) measure $\mu$,[6–8]

$$\mathbb{E}[a] = \lim_{T\to\infty} \frac{1}{T}\int_0^T a(\mathbf{x}(t))\,dt \tag{1}$$

$$\equiv \int_{\mathscr{X}} a(\mathbf{x})\,d\mu(\mathbf{x}), \tag{2}$$

for all $a(\mathbf{x})$ that are $\mu$-integrable. Here, $\mathscr{X}$ denotes an ergodic subset of the state space $\mathscr{M}$, and $\mathbf{x}(t)$ is a trajectory confined to, or

approaching, $\mathscr{X}$. Note that $\mathbb{E}[a]$ is invariant with respect to the choice of $\mathbf{x}(0) \in \mathscr{X}$.

Evaluating $\mathbb{E}[\cdot]$ exactly requires an infinitely long trajectory $\mathbf{x}(t)$ that explores $\mathscr{X}$ or the knowledge of $\mu$ itself, either of which is only available in highly exceptional cases.[9] By necessity, $\mathbb{E}[\cdot]$ is commonly expressed as the average of $a(\mathbf{x}(t))$ over a finite but sufficiently long trajectory. In high-dimensional systems, this trajectory-based full order model of $\mathbb{E}$ becomes exceedingly expensive and cumbersome: for instance, a turbulent fluid flow on a relatively small computational domain and over a relatively short time interval requires hundreds of terabytes of storage.[10]

A more practical, and far more data-efficient, alternative is to define a collection $\{\mu_p\}_{p=1}^{P}$ of measures supported on subsets $S_p \equiv \operatorname{supp} \mu_p \subseteq \mathscr{M}$ such that

$$\hat{\mathbb{E}}[a] \equiv \sum_{p=1}^{P} w_p \, \mathbb{E}_p[a] \approxeq \mathbb{E}[a] \tag{3}$$

is a good approximation of $\mathbb{E}[a]$, where

$$\mathbb{E}_p[a] = \int_{S_p} a(\mathbf{x}) d\mu_p(\mathbf{x}) \tag{4}$$

is the integral of $a$ over $\mu_p$, and $w_p$ are scalar weights. For finite $P$, Eq. (3) provides a reduced order model for evaluating time averages of the dynamics. When the weights $\{w_p\}_{p=1}^{P}$ are non-negative and sum to one, Eq. (3) can be interpreted as a non-Gaussian mixture model of a system's stationary statistics. Various expansions of this form have been proposed,[11–13] with different choices of $S_p$, means for computing $w_p$, advantages and drawbacks.

Equation (3) can only be used to predict the averages of observables for which $\mathbb{E}[a]$ and $\mathbb{E}_p[a]$ exist and are finite. It is common to require that all measures are normalizable such that $0 < |\mathbb{E}[1]|, |\mathbb{E}_p[1]| < \infty$. This condition ensures that $\mathbb{E}[a]$ and $\mathbb{E}_p[a]$ are finite for all bounded observables, i.e., those satisfying $\sup_{\mathbf{x}\in\mathscr{M}} |a(\mathbf{x})| < \infty$. Without loss of generality, it is further assumed that all averages are normalized, $\mathbb{E}[1] = \mathbb{E}_p[1] = 1$.

Statistical reduced order models like Eq. (3) are particularly well-suited for computing the averages that depend on hyper-parameters. Examples include the $\gamma$-norm of the state vector, $a_\gamma(\mathbf{x}) = \left(\sum_i x_i^\gamma\right)^{-\gamma}$, the variance of the flow about a reference state, $a_{\boldsymbol{\gamma}}(\mathbf{x}) = (\mathbf{x}-\boldsymbol{\gamma})^\top(\mathbf{x}-\boldsymbol{\gamma})$, or even the Fourier transform of the SRB measure, $a_{\boldsymbol{\gamma}}(\mathbf{x}) = e^{i\boldsymbol{\gamma}\cdot\mathbf{x}}$. The nonlinear dependence of $a_{\boldsymbol{\gamma}}$ on $\boldsymbol{\gamma}$ renders direct computation of the averages of such an observable over a range of $\gamma$ very expensive. Traditionally, the hyper-parameter space must be discretized, and an infinite-time average $\mathbb{E}[a_{\boldsymbol{\gamma}}]$ computed for each $\boldsymbol{\gamma}$ on a grid. Equation (3) allows for, at every grid point, the expensive average over the chaotic trajectory to be replaced with inexpensive averages over reference states. For fine discretizations of $\boldsymbol{\gamma}$, this can result in massive speedup.

A particularly desirable choice for the subsets $S_p$ is the unstable periodic orbits (UPOs).[14–16] Periodic orbits are special, repeating solutions of a dynamical system satisfying $\mathbf{x}_p(t) = \mathbf{x}_p(t+T_p)$, where $T_p$ is the period of motion. Due to periodicity, averages over UPOs can be computed exactly in finite time,

$$\mathbb{E}_p[a] = \frac{1}{T_p}\int_0^{T_p} a(\mathbf{x}_p(t))\, dt, \tag{5}$$

allowing the infinite-time average present in Eq. (3) to be replaced with a collection of finite time averages.

This choice, which is the basis of the periodic orbit theory (POT), leverages the fact that periodic orbits are dense in the closure of the ergodic set of Axiom A systems.[17,18] In particular, for Axiom A systems, it can be proven that Eq. (3) converges to equality for smooth observables[12] in the limit $P \to \infty$. Moreover, the weights $w_p$ may be computed analytically in terms of the stability properties of the UPOs.[19] A straightforward generalization of POT exists for systems with continuous symmetries, where UPOs become relative;[20] we will refer to both types of solutions as simply UPOs.

Periodic orbits often exemplify the same spatiotemporal features as the chaotic flow itself[21–24] and are considered to form a dynamical skeleton of $\mathscr{X}$.[25–28] Hence, when periodic orbits are known, they generate a remarkably computationally efficient and *interpretable* representation of the chaotic averages, particularly in high dimensions. For this reason, practitioners have sought, and struggled, to apply POT to high-dimensional chaotic flows for more than two decades.[12,29–33]

It is notoriously difficult to compute a sufficiently comprehensive collection of periodic orbits in high dimensions.[30] For instance, in turbulent fluid flows, the number of known periodic orbits rarely exceeds a few tens.[25,26,30,31,34,35] Furthermore, the orbits that are discovered tend to sample the chaotic set poorly, a limitation that only state-of-the-art methods have begun to address.[32]

Also, the convergence of POT relies critically on the existence of a *symbolic dynamics*. Symbolic dynamics is a mathematical framework that simplifies the description of a system by partitioning its state space into distinct regions and representing trajectories by the sequence of regions they visit.[36] In ergodic systems, this partition is meticulously constructed to reflect the underlying topology of the chaotic set. For periodic orbits, which repeat in time, their symbolic sequences are likewise periodic, with the symbol length defined as the number of symbols before the sequence repeats. A library of orbits is said to be complete up to symbol length $L$ if it includes—and only includes—all periodic orbits with symbol length $l < L$. POT is proven to converge exponentially fast for successively *complete* libraries, in the limit $L \to \infty$. However, POT is only as accurate as the shortest orbit truncated from the expansion; if a short orbit is missed, adding more orbits only negligibly improves the accuracy of the method.[37]

It is important to note that most systems of practical importance do not admit or do not have known symbolic dynamics, rendering it impossible to tell when a library is complete and further complicating the usage of POT. Stability ordering—an extension of POT to incomplete libraries and systems without symbolic dynamics—has been proposed (see Ch. 23.7 of Ref. 12). However, this is an active area of research and the expected accuracy of stability ordering has not yet been established. These issues, collectively, have prevented POT from producing accurate predictions in high-dimensional systems.[30]

A popular alternative to POT is Ulam's method,[11,38] which typically partitions a compact set enclosing $\mathscr{X}$ into disjoint subsets $\{S_p\}_{p=1}^{P}$, each imbued with a uniform measure. This approach is proven to approximate averages over any absolutely continuous measure in the limit that the size of the subsets becomes vanishingly small. However, this method also struggles in applications to high-dimensional flows. Even for a moderate number of dimensions, Ulam's method quickly becomes intractable due to the proliferation of sets and difficulty associated with integrating over the uniform measure in $d$ dimensions. Some studies are attempting to address these issues with novel clustering methods.[13]

Lacking many prerequisites for a working periodic orbit theory, practitioners working in high-dimensional systems have pursued a periodic orbit decomposition of chaotic averages in which—in lieu of the POT weights—$w_p$ are computed from a Markov process defined on the Voronoi volumes about each UPO.[32,33] Empirically, Markov-based approaches achieve moderate accuracy on the order of a few percent. However, the Markov-based approach lacks theoretical guarantees regarding the types of observables that can be accurately predicted using small sets of UPOs and whether it converges in the limit $P \to \infty$.

In this letter, we introduce an interpretable data-driven approach to determining optimal weights for a finite, and possibly quite small, library of measures $\{\mu_p\}_{p=1}^{P}$. It utilizes Kriging,[39] a method that leverages a Reproducing Kernel Hilbert Space (RKHS) to perform interpolation and regression on abstract data. Kriging is known for its accuracy in applications to high-dimensional data sets.[40–42] The method enables averages over a target measure $\mu$ to be estimated from *any* choice of reference measures $\{\mu_p\}_{p=1}^{P}$. It also provides rigorous claims as to which observables averages can be computed with zero error. The proposed method finds multiple distributions of weights that are more accurate and converge more rapidly in $P$ than both POT and Markov models.

## II. THE METHOD

The method being proposed here is quite simple. To illustrate it, we begin with a tutorial before presenting a more rigorous derivation.

### A. Tutorial

Suppose that $P = 2$ and that $S_p = \{\mathbf{x}_p(t) \mid 0 \le t < T_p\}$ where $p = 1, 2$ are sets associated with periodic orbits. Then, Eq. (3) reduces to

$$\mathbb{E}[a] \approxeq w_1 \mathbb{E}_1[a] + w_2 \mathbb{E}_2[a]. \tag{6}$$

The above equation is linear, with two unknowns $w_1$ and $w_2$. By evaluating Eq. (6) on two or more test observables, we may solve for the weights using the method of least squares.

For instance, suppose that $a_1(\mathbf{x})$ and $a_2(\mathbf{x})$ are observables of particular interest. Then, weights that best predict the chaotic averages of $a_1$ and $a_1$ from averages over orbits 1 and 2 will approximately satisfy Eq. (6) for both $a_1$ and $a_2$,

$$\begin{pmatrix} \mathbb{E}[a_1] \\ \mathbb{E}[a_2] \end{pmatrix} \approx \begin{pmatrix} \mathbb{E}_1[a_1] & \mathbb{E}_2[a_1] \\ \mathbb{E}_1[a_2] & \mathbb{E}_2[a_2] \end{pmatrix} \begin{pmatrix} w_1 \\ w_2 \end{pmatrix}. \tag{7}$$

A least squares solution

$$\begin{pmatrix} w_1 \\ w_2 \end{pmatrix} = \begin{pmatrix} \mathbb{E}_1[a_1] & \mathbb{E}_2[a_1] \\ \mathbb{E}_1[a_2] & \mathbb{E}_2[a_2] \end{pmatrix}^{\dagger} \begin{pmatrix} \mathbb{E}[a_1] \\ \mathbb{E}[a_2] \end{pmatrix} \tag{8}$$

always exists, where $[\,\cdot\,]^{\dagger}$ indicates a pseudo-inverse. Solving Eq. (7) when the matrix is non-invertible is further discussed in Sec. IV.

As written, the weights given in Eq. (8) are heavily biased towards predicting well the averages of $a_1$ and $a_2$. They may predict other observables' averages arbitrarily poorly. To avoid biasing the least squares solution toward any particular physical observable, consider instead a spatially localized observable, such as a Gaussian kernel

$$G_\theta(\mathbf{x}, \mathbf{y}) = \exp\left[-\frac{|\mathbf{y} - \mathbf{x}|^2}{2\theta}\right], \tag{9}$$

with variance $\theta$ and mean $\mathbf{y}$. Evaluating Eq. (6) on $G_\theta$, it follows that

$$f(\mathbf{y}) \approxeq w_1 f_1(\mathbf{y}) + w_2 f_2(\mathbf{y}), \tag{10}$$

where

$$f(\mathbf{y}) = \mathbb{E}[G_\theta(\,\cdot\,, \mathbf{y})] \tag{11}$$

$$= \int_{\mathscr{M}} G_\theta(\mathbf{x}, \mathbf{y}) d\mu(\mathbf{x}) \tag{12}$$

$$= \lim_{T\to\infty} \frac{1}{T} \int_0^T G_\theta(\mathbf{x}(t), \mathbf{y}) dt \tag{13}$$

is a function related to the probability that the chaotic trajectory visits the $\theta$-neighborhood of $\mathbf{y}$, and

$$f_p(\mathbf{y}) = \mathbb{E}_p[G_\theta(\,\cdot\,, \mathbf{y})] \tag{14}$$

$$= \int_{S_p} G_\theta(\mathbf{x}, \mathbf{y}) d\mu_p(\mathbf{x}) \tag{15}$$

$$= \frac{1}{T_p} \int_0^{T_p} G_\theta(\mathbf{x}_p(t), \mathbf{y}) dt \tag{16}$$

is a function related to the probability that orbit $p$ visits the $\theta$-neighborhood of $\mathbf{y}$. In the limit that $\theta$ becomes small, $f(\mathbf{x})$ limits to—up to normalization—the probability that a randomly sampled chaotic state is $\mathbf{x}$. Indeed, $\theta$ plays a regularizing role quite similar to that of noise; it has been shown previously that—in the presence of noise—this probability can be expanded as a linear sum of Gaussians supported on periodic points in the chaotic set.[43] Equation (10) expresses a similar idea. Examples of $f(\mathbf{x})$ and $f_p(\mathbf{x})$ are illustrated in Fig. 1 for the Lorenz 1963 system[4] with canonical parameter values $\sigma = 10$, $\rho = 28$, and $\beta = 8/3$; Lorenz is a reduced order model for atmospheric convection and a prototypical example of deterministic chaos.

Since all choices of $\mathbf{y} \in \mathscr{M}$ constrain the weights, Eq. (10) represents an *over-constrained* system of equations for weights $w_1$ and $w_2$. It can also be solved using the method of least squares. The resulting optimal weights are as follows:

$$\begin{pmatrix} w_1 \\ w_2 \end{pmatrix} = \begin{pmatrix} \langle f_1, f_1 \rangle & \langle f_1, f_2 \rangle \\ \langle f_2, f_1 \rangle & \langle f_2, f_2 \rangle \end{pmatrix}^{\dagger} \begin{pmatrix} \langle f_1, f \rangle \\ \langle f_2, f \rangle \end{pmatrix}, \tag{17}$$

where

$$\langle f, g \rangle = \int_{\mathscr{M}} f(\mathbf{x}) g(\mathbf{x}) d\mathbf{x} \tag{18}$$

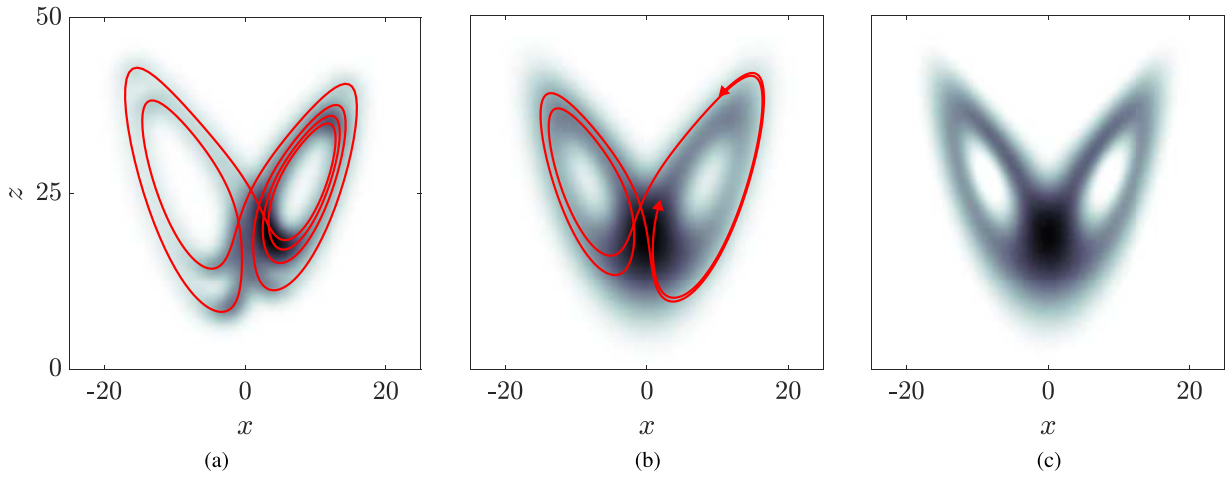


**FIG. 1.** An illustration of (a) periodic orbit density $f_p(\mathbf{x})$ and (b) chaotic orbit density $f(\mathbf{x})$ over the Lorenz 1963 dynamics using a Gaussian kernel with unit variance. In each, the trajectory that generates the data is shown in red. Darker colors indicate larger values of $f$. Red arrows illustrate that the chaotic trajectory extends infinitely in both directions, without repeating. (c) A histogram of all states visited by a long chaotic trajectory, color indicated the probability that a bin is visited by the chaotic trajectory. Darker bins are visited more often. In all panels, the function (or histogram) is averaged over $y$ for plotting purposes.

is an inner product on $\mathscr{M}$.

Integration over a state space—as required by Eq. (18)—is likely numerically intractable. However, a well-chosen kernel will allow the integral over $\mathscr{M}$ to be taken analytically. Notice that for a Gaussian kernel in $\mathscr{M} = \mathbb{R}^d$, only integrals over orbits $p$ and $q$ are required,

$$\langle f_p, f_q \rangle = \int_{\mathscr{M}} f_p(\mathbf{x}) f_q(\mathbf{x})\, d\mathbf{x} \tag{19}$$

$$= \int_{S_p} d\mu_p(\mathbf{y}) \int_{S_q} d\mu_q(\mathbf{z}) \int_{\mathscr{M}} G_\theta(\mathbf{y},\mathbf{x}) G_\theta(\mathbf{x},\mathbf{z})\, d\mathbf{x}$$

$$= (\pi\theta)^{d/2} \int_{S_p} d\mu_p(\mathbf{y}) \int_{S_q} d\mu_q(\mathbf{z})\, G_{2\theta}(\mathbf{y},\mathbf{z}) \tag{20}$$

$$= \frac{(\pi\theta)^{d/2}}{T_p T_q} \int_0^{T_p} \int_0^{T_q} G_{2\theta}(\mathbf{x}_p(t), \mathbf{x}_q(\tau))\, dt d\tau. \tag{21}$$

Hence, these inner products can be computed as averages over orbits, which are computationally inexpensive.

Unlike the weights given in Eq. (8), the solution provided in Eq. (17) is unbiased toward any physical observable. Rather, these weights optimize the global reconstruction of $f(\mathbf{x})$ from functions $f_p(\mathbf{x})$.

In Sec. II B, we re-derive the method for an arbitrary kernel function observable and define the subspace of observables $\mathscr{A}$ whose averages are predicted by the method with zero error.

### B. General approach

Consider a smooth, bounded kernel function $k(\mathbf{x},\mathbf{y}) \in L^2(\mathscr{M} \times \mathscr{M})$ satisfying

$$\int_{\mathscr{M}} k(\mathbf{x},\mathbf{z}) k(\mathbf{z},\mathbf{y})\, d\mathbf{z} = k^2(\mathbf{x},\mathbf{y}), \tag{22}$$

where $k^2$ is an arbitrary induced kernel function. A particularly convenient choice of $k$ is $k(\mathbf{x},\mathbf{y}) = k_\theta(\|\mathbf{x}-\mathbf{y}\|)$, where $k_\theta$ is part of a parametric family of functions closed under convolution. Both Gaussian and Cauchy distributions are of this type. In fact, all stable distributions are of this type. Because $k$ is smooth and bounded, so is $k^2$, and both are guaranteed to be $\mu$- and $\mu_p$-integrable. This allows us to define the correlation between arbitrary normalizable measures $\mu'$ and $\mu''$ (supported on sets $S', S'' \subset \mathscr{M}$) as

$$K(\mu',\mu'') \equiv \int_{\mathscr{M}} \mathbb{E}_{\mu'}\left[k(\cdot,\mathbf{z})\right] \mathbb{E}_{\mu''}\left[k(\cdot,\mathbf{z})\right]\, d\mathbf{z}$$

$$= \int_{S'} d\mu'(\mathbf{x}) \int_{S''} d\mu''(\mathbf{y})\, k^2(\mathbf{x},\mathbf{y}). \tag{23}$$

Intuitively, one can think of $K(\mu',\mu'')$ as an inner product of measures $\mu'$ and $\mu''$. Indeed, when $k$ is a Gaussian kernel function, $K(\mu_p,\mu_q) = \langle f_p, f_q \rangle$ is the inner product from Eq. (20).

Let us associate a unique observable $a_p(\mathbf{x})$ with each orbit measure $\mu_p$ by defining

$$a_p(\mathbf{x}) \equiv \int_{S_p} d\mu_p(\mathbf{y})\, k^2(\mathbf{y},\mathbf{x}). \tag{24}$$

Let $\mathscr{A} = \text{span}\,(a_1(\mathbf{x}), \ldots, a_P(\mathbf{x}))$ be the span of these functions. For any observable $a \in \mathscr{A}$, i.e.,

$$a(\mathbf{x}) = \sum_{p=1}^{P} c_p a_p(\mathbf{x}), \tag{25}$$

using Eq. (23) we find that

$$\int d\mu'(\mathbf{x})\, a(\mathbf{x}) = \sum_p c_p K(\mu_p,\mu'). \tag{26}$$

Evaluating Eq. (26) at $\mu' = \mu_q$ for $q = 1, \ldots, P$, we find

$$\sum_{p=1}^{P} c_p A_{pq} = \mathbb{E}_q[a], \tag{27}$$

where $A_{pq} = K(\mu_p, \mu_q)$.

Define weights $w_p$ as the solution to the system

$$\sum_{q=1}^{P} A_{pq} w_q = \mathbb{E}[a_p]. \tag{28}$$

Then, averaging both sides of Eq. (25) over the SRB measure, it follows that

$$\mathbb{E}[a] = \sum_{p=1}^{P} c_p \mathbb{E}[a_p] = \sum_{q=1}^{P} w_q \mathbb{E}_q[a] = \hat{\mathbb{E}}[a], \tag{29}$$

where the equality between $\mathbb{E}[a]$ and $\hat{\mathbb{E}}[a]$ is *exact*, rather than approximate, for all $a \in \mathscr{A}$.

Hence, Eq. (28) recasts computing optimal weights as solving a simple linear system of the form $A\vec{w} = \vec{b}$, where $b_p = \mathbb{E}[a_p]$ is a vector of $P$ time averages. We call $\hat{\mathbb{E}}[a]$ with the weights given by Eq. (28) a least squares weighted (LSW) estimate of $\mathbb{E}[a]$.

For observables $a \notin \mathscr{A}$, using the least squares weights to compute $\hat{\mathbb{E}}[a]$ may not yield an exact value for $\mathbb{E}[a]$. However, the error can remain quite small for observables outside $\mathscr{A}$. For instance, since the observables in $\mathscr{A}$ are smooth functions of the state $\mathbf{x}$ (as a consequence of smoothness of the kernel), one can assume that low error can be similarly achieved for smooth observables. We indeed find this to be the case, as evidenced in Sec. III.

## III. RESULTS

We compare the accuracy of different approaches using the Lorenz 1963 system. Lorenz admits a symbolic dynamics, making it an ideal testbed in which to compare LSW against POT. For an introduction to symbolic dynamics in Lorenz, we point the reader to Viswanath.[44] In this system, an orbit's symbol length is equal to the number of times it winds around the left and right sides of chaotic set, before repeating; see Fig. 1(a) for an example of a UPO with symbol length $l = 6$. We will use the $P \leq 125$ UPOs of the shortest symbol length from a collection of more than a thousand found by Viswanath.[44,45] To ensure a fair and rigorous comparison between LSW and POT, we compute POT weights exclusively for complete libraries of symbol length $l = 2, \ldots, 9$. The sizes of these complete libraries are $P = 1, 3, 6, 12, 21, 39, 69$, and 125, respectively. As previously stated, we are particularly interested in computing averages when $P$ is small, which will be the most realistic regime in applications to higher dimensional chaotic systems.

In addition to computing averages for periodic orbits, we will also consider a collection of short, finite-time intervals sampled from chaotic trajectory (which we call "snippets"). Computing periodic orbits requires expensive numerical solvers. For instance, even for a strongly confined Taylor–Couette flow at moderate Reynolds numbers, it takes tens of hours to converge a single relative periodic orbit on a typical workstation.[26] In contrast, sampling snippets from numerical integration is trivial; even computing snippets that are *almost*-periodic from auto-recurrence is comparatively inexpensive. Snippets allow us to explore the applicability of a LSW estimate to systems for which no, or too few, UPOs are available.

To generate snippets, we compute a chaotic trajectory that explores $\mathscr{X}$ for $\sum_{p=1}^{125} T_p$ time units and, for simplicity, divide it uniformly into 125 snippets of equal duration. In this way, the total length of all the UPOs and all snippets considered is the same. We do not ensure that the snippets are well-distributed on the attractor or enforce upon the collection of snippets any symmetries of the governing equations.

We use a Gaussian kernel function $k = G_\theta$, as introduced in the tutorial. For $\theta$ too large or too small, the correlation matrix $A$ reduces to a matrix of all ones or to a diagonal matrix, respectively; in the former limit, the matrix $A$ becomes entirely singular and, in the latter, the LSW weights limit to Markovian-like weights defined on sets $S_p$ (that is, $w_p \to \mathbb{E}\left[I(\mathbf{x}, S_p)\right]$, where $I(\mathbf{x}, S_p)$ is an indicator function equal to unity when $\mathbf{x} \in S_p$ and zero otherwise). Both extremes wash away the rich correlations between measures $\mu_p$. We use $\theta = 10^2$ which, consistently, across both UPOs and snippets, is roughly equidistant—in terms of the Frobenius norm—from either extreme, as illustrated in Fig. 2. For reference, the corresponding value of $\sqrt{\theta}$ is of order $\sqrt{2\beta(\rho - 1)} = 12$, which sets the size of the attractor. There are examples of more complex choices of $\theta$ in the literature; for instance, Lippolis and Cvitanović[46] define a spatially varying Gaussian kernel width, proportional to the local response of a dynamics to noise. Here, we find a global choice of $\theta$ based on the Frobenius distance to be simple yet efficient.

For any fixed $\theta$, $A$ is likely to become increasingly singular as $P$ increases, due to increasing overlap between the functions $a_p$. To penalize marginal modes from appearing in the solution weights, we solve Eq. (28) using Tikhonov regularization,

$$(A + \alpha I)\, \vec{w} = \vec{b}, \tag{30}$$

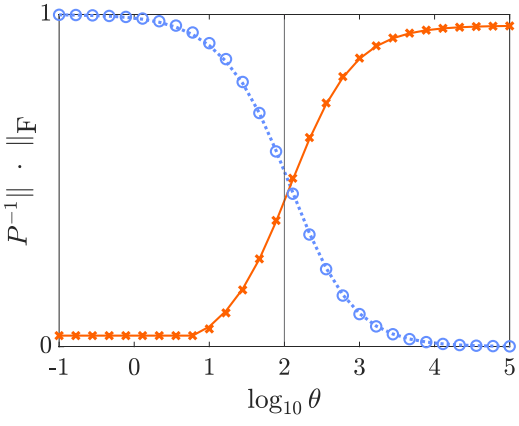


**FIG. 2.** The Frobenius distance between $A$ and the matrix of all ones (blue, dotted line, o's) or the identity matrix (orange, solid line, x's) for $P = 30$. Lines (symbols) correspond to matrices $A$ computed using UPOs (snippets). Black line indicates the value of $\theta = 10^2$ used in Fig. 3 and Table I.

with $\alpha = 10^{-10}$. The marginal modes of $A$ are discussed further in Sec. IV.

We compute the elements of matrix $A$ and $\vec{b}$ using the formulas

$$A_{pq} = \int_0^{T_p} \frac{dt}{T_p} \int_0^{T_q} \frac{d\tau}{T_q} \exp\left[-\frac{\|\mathbf{x}_p(t) - \mathbf{x}_q(\tau)\|^2}{4\theta}\right], \quad (31)$$

$$b_q = \lim_{T\to\infty} \int_0^{T} \frac{dt}{T} \int_0^{T_q} \frac{d\tau}{T_q} \exp\left[-\frac{\|\mathbf{x}(t) - \mathbf{x}_q(\tau)\|^2}{4\theta}\right], \quad (32)$$

The largest computational cost associated with this method is the evaluation of the averages in $\vec{b}$. However, these averages are independent and can be computed in parallel (across orbits and in time) quite quickly, once and for all, no matter how many different averages we might need to estimate later.

We cross-validate the accuracy of LSW on a set of observables not in $\mathscr{A}$, namely, $\mathscr{B} = \{1, x, y, z, x^2, xy, xz, y^2, yz, z^2\}$. All second order polynomials, including observables like $\dot{\mathbf{x}}$, lie in the span of the elements of $\mathscr{B}$. We define the relative error

$$E_{\text{rel}}(a) = \frac{\left|\mathbb{E}[a] - \hat{\mathbb{E}}[a]\right|}{\sqrt{\text{var}(a)}} \quad (33)$$

and maximal error

$$E_{\text{max}} = \max_{b\in\mathscr{B}} E_{\text{rel}}(b), \quad (34)$$

where the variance, $\text{var}(a)$, sets a natural scale for deviations from $\mathbb{E}[a]$. For observables with zero variance, we set $\text{var}(a) \equiv 1$.

Additionally, we use LSW to estimate global properties of the chaotic set. In particular, it is possible to construct estimates of the Lyapunov exponents, $\lambda^i$, from weighted averages over orbits (see Appendix B). When sufficiently many Lyapunov exponents of a system are known (or estimated), one can also compute (or estimate) the Kaplan–Yorke dimension, $d_{KY}$, of the chaotic set (see Appendix B). Since these quantities do not have a well-defined variance over the chaotic set, we simply report the absolute error of their estimates.

We compare the accuracy of the proposed approach against three alternatives. The first one is POT, which has been applied to Lorenz previously and shown to provide accurate estimates of averages for a variety of observables.[14] The second is a Markov weighting. Lastly, we consider a uniform weighting, where $w_p = 1/P$. Uniform weighting is the simplest option which has been used previously as a baseline with which to compare other weighting schemes.[30] While POT can only be applied to UPOs, all other methods can be applied to snippets as well.

There are four parameters that can impact the weights. There is (i) the cardinality of the library, $P$, as well as (ii) the choice of *which* $P$ measures are used. To explore this, we randomly permute the entire collection of 125 measures (orbits or snippets) $R - 1 = 255$ different times, labeling these libraries $\{\mathscr{L}_r\}_{r=2}^{R=256}$; we retain $\mathscr{L}_1$ as an un-permuted library, in which UPOs are ordered by increasing symbol length. This results in $R = 256$ libraries total. At each value of $P$, weights are computed using the first $P$ measures of $\mathscr{L}_r$, for each $r$. The LSW and Markov weights, which are computed from chaotic data, can also depend on how that data are collected. Namely, (iii) where the chaotic trajectory is initialized and (iv) how long the chaotic trajectory is or, more specifically, how many snapshots from the chaotic trajectory are used. To explore this, we compute 256 different chaotic trajectories, $\{\mathbf{x}^{(s)}(n\Delta t)\}_{s=1}^{256}$, where $\Delta t = 2$ and $n = 1, \ldots, 10^6$. Each chaotic sample is randomly initialized at time $t = -25$. This random initial condition is then integrated to time $t = 0$ and only data with $t > 0$ are retained. This ensures all trajectories are randomly initialized at time $t = 0$ within the chaotic set. Averages required for calculating the LSW or Markov weights are then calculated, for each $s$, using only the first $N$ samples of the chaotic trajectory. Specifically, the LSW weights are computed via Eq. (28) with

$$\mathbb{E}[a_p] \approx \frac{1}{N}\sum_{n=1}^{N} a_p\left(\mathbf{x}^{(s)}(n\Delta t)\right). \quad (35)$$

Similarly, the Markov weights are computed as

$$w_p = \mathbb{E}\left[I(\mathbf{x}, V_p)\right] \approx \frac{1}{N}\sum_{n=1}^{N} I\left(\mathbf{x}^{(s)}(n\Delta t), V_p\right), \quad (36)$$

where the Voronoi volume $V_p$ is defined as the collection of points $\mathbf{x} \in \mathscr{M}$ closer to subset $S_p$ than any other subset,

$$V_p = \left\{\mathbf{x} \in \mathscr{M} \mid arg\,min_{q\in\mathscr{L}_r} \min_{\mathbf{y}\in S_q} \|\mathbf{x} - \mathbf{y}\| = p\right\}, \quad (37)$$

and $I$ is the previously mentioned indicator function. Hence, for Markov weights, $w_p$ is simply the fraction of time $\mathbf{x}(t)$ is closer to subset $S_p$ than any other subset. Sampling over $P$, $r$, $s$, and $N$ allows us to eliminate most of the bias and to quantify the uncertainty in the error associated with each of the four weighting schemes. Of these four degrees of freedom, the uniform weighting scheme will vary only with $P$ and $r$, since it does not require chaotic data. Similarly, POT weights will vary only with $P$, as they are only computed for complete libraries (i.e., $r = 1$ for specific $P$).

Table I summarizes $E_{\text{rel}}$ for each method and each observable in $\mathscr{B}$ at a representative choice of $P$ and $N$. Table I also lists the absolute error associated with estimating $\lambda^1$, $\lambda^3$, and $d_{KY}$ (the value of $\lambda^2$ is predicted with zero error by all weighting schemes). We find that, excluding errors that are identically zero, the two most accurate estimators of chaotic averages, across all observables, are the LSW estimators over UPOs and snippets. The LSW estimator over orbits almost always outperforms the estimator over snippets and consistently outperforms competing methods by nearly an order of magnitude.

For a number of observables, the error associated with POT is identically zero. This is because the Lorenz attractor is invariant under the transformation $(x, y, z) \mapsto (-x, -y, z)$. As a result, the infinite time average of any monomial that is odd in $x$ or $y$ is zero, e.g., $\mathbb{E}[x] = \mathbb{E}[y] = \mathbb{E}[xz] = \mathbb{E}[yz] = 0$. For every orbit $p$ of the Lorenz system, there exists a corresponding orbit $q$ such that $(x_p(t), y_p(t), z_p(t)) = (-x_q(t), -y_q(t), z_q(t))$. Since these orbits must have identical stability properties, POT will assign them the same weight, $w_p = w_q$. Yet, they will produce equal but opposite time averages of any monomial that is odd in $x$ or $y$, $\mathbb{E}_p[a] = -\mathbb{E}_q[a]$. Since $p$ and $q$ must have the same symbol length, every *complete* library is guaranteed to contain both orbits and their contributions

**TABLE I.** The median value of $\log_{10}(E_{\text{rel}})$ over $s$ and $r$ for each weighting scheme and each observable in $\mathscr{B}$ at $P = 21$ and $N = 10^6$. POT is only evaluated at $r = 1$. The last three rows report $\log_{10}$ the absolute error of predicting Lyapunov exponents and Kaplan–Yorke dimension from periodic orbits. Column headers are color-coded to match Fig. 3.

| | Orbits | | | | Snippets | | |
|---|---|---|---|---|---|---|---|
| | POT | Uniform | Markov | LSW | Uniform | Markov | LSW |
| 1 | $-\infty$ | $-\infty$ | $-\infty$ | $-3.9$ | $-\infty$ | $-\infty$ | $-4.4$ |
| $x$ | $-\infty$ | $-1.4$ | $-1.7$ | $-3.2$ | $-1.2$ | $-1.6$ | $-3.1$ |
| $y$ | $-\infty$ | $-1.4$ | $-1.7$ | $-3.3$ | $-1.3$ | $-1.6$ | $-3.1$ |
| $z$ | $-1.7$ | $-1.8$ | $-2.0$ | $-3.5$ | $-2.1$ | $-2.2$ | $-3.3$ |
| $x^2$ | $-2.2$ | $-2.3$ | $-2.5$ | $-4.0$ | $-2.4$ | $-2.4$ | $-3.3$ |
| $xy$ | $-2.3$ | $-2.3$ | $-2.6$ | $-4.0$ | $-2.4$ | $-2.4$ | $-3.2$ |
| $xz$ | $-\infty$ | $-1.4$ | $-1.7$ | $-3.3$ | $-1.3$ | $-1.6$ | $-3.1$ |
| $y^2$ | $-2.1$ | $-2.2$ | $-2.4$ | $-4.0$ | $-2.4$ | $-2.4$ | $-3.2$ |
| $yz$ | $-\infty$ | $-1.4$ | $-1.7$ | $-3.3$ | $-1.3$ | $-1.6$ | $-3.1$ |
| $z^2$ | $-1.9$ | $-2.0$ | $-2.3$ | $-3.7$ | $-2.3$ | $-2.3$ | $-3.3$ |
| $\lambda^1$ | $-1.5$ | $-1.6$ | $-1.9$ | $-3.1$ | | | |
| $\lambda^3$ | $-1.5$ | $-1.6$ | $-1.9$ | $-2.7$ | | | |
| $d_{KY}$ | $-2.7$ | $-2.8$ | $-3.1$ | $-4.3$ | | | |

must cancel, $w_p\mathbb{E}_p[a] + w_q\mathbb{E}_q[a] = 0$. As a result, the POT weighting also predicts a zero infinite time average for any monomial that is odd in $x$ or $y$. It is important to note that this property of POT is incredibly sensitive to the library being complete. If a single orbit is missed or an extraneous orbit included, this property no longer holds.

For $a = 1$, the errors associated with all non-LSW weightings are identically zero. This is because each of these methods returns normalized weights, $\sum_{p=1}^{P} w_p = 1$, by construction; an accurate prediction of $a = 1$ relies only on the weights being normalized, $1 = \mathbb{E}[1] = \sum_{p=1}^{P} \mathbb{E}_p[1]w_p = \sum_{p=1}^{P} w_p$. In contrast, LSW solutions are not constrained to be normalized. Yet, LSW *discovers* solutions that are normalized to four significant digits, see Table I. Notably, normalization of the LSW solution can be enforced to machine precision, if desired, which we explore in Sec. IV.

Figure 3 shows the dependence of $E_{\max}$ on $N$ and $P$. At fixed $N$, $E_{\max}$ decreases more rapidly in $P$ for the LSW estimate than for any other method, see Figs. 3(b) and 3(d); this is the case for both orbits and snippets. For $N = 10^6$, the LSW estimate is often an order of magnitude more accurate than the other methods. Most importantly, LSW weights yield an accurate estimate for all considered observables from as few as 20 orbits or snippets. This corresponds to accurately reproducing averages computed directly via Eq. (35) over $T = 256 \times N\Delta t = O(10^9)$ time units worth of chaotic data using only $T = O(10^2)$ time units worth of orbit or snippet data.

As illustrated by Figs. 3(a) and 3(c), the error associated with the LSW estimate scales as $E_{\max} \sim N^{-1/2}$ for sufficiently low $N$ and high $P$. This is most likely due to the accuracy of approximating $\mathbb{E}[a_p]$ via Eq. (35) determined by the central limit theorem. The Markov error also scales like $E_{\max} \sim N^{-1/2}$ and often slightly outperforms the LSW estimate, when $N < 10^3$. As $N$ increases above $10^3$, the Markov estimator converges to about 1% error. Hence, when chaotic data are very sparse, it seems well-motivated to use a Markov estimator of infinite-time averages. When more data are available, LSW provides noticeably more accurate estimates.

To the authors' knowledge, the convergence rates of the Markov, Uniform, and POT methods have not been directly compared in prior studies. Surprisingly, our results show that, for the range of $P$ considered here, POT does not yield meaningfully better estimates than the empirical approaches. The convergence rates, in the number of orbits/snippets, of all non-LSW orbit expansions are largely comparable. Moreover, the error of Markov weights appears to converge faster than that of POT for moderate values of $P$ considered here. When applied to snippets, Markov weights consistently outperform uniform weights; both weighting schemes appear to asymptote at large values of $P$.

A notable advantage of POT remains that its errors do not plateau as a function of $P$; unlike the error of empirical approaches, the error of POT is proven to converge to zero as the library of orbits used becomes increasingly large and complete. On the other hand, the LSW method is able to provide an estimate of its asymptotic accuracy, $E_{\max} \sim N^{-1/2}$, a capability absent in both Markov and uniform weightings. For Markov weighting, it is the variance of the weights which scales like $N^{-1/2}$, rather than the accuracy of the estimate.

## IV. UNIQUENESS OF LEAST SQUARES WEIGHTS

The linear Eq. (28) has a unique solution when the matrix $A$ is non-singular, e.g., in the limit $\theta \to 0$. Convolution with a Gaussian kernel is a smoothing operation that introduces increasingly many marginal directions into $A$ as $\theta$, the smoothing parameter, is increased. In the limit $\theta \to \infty$, Eq. (10) limits to a single normalization constraint, $1 = \sum_p w_p$, that is independent of $\mathbf{y}$; in this limit, the weights become severely under-constrained. Hence, Eq. (28) generally admits multiple solutions.

Marginal modes that are introduced at finite values of $\theta$ are rather interesting. Each marginal mode of $A$ approximately satisfies $\sum_p w_p = 0$ and so generates a splitting of the set of orbits into two subsets, one with positive weights $\mathscr{P} = \{p|w_p > 0\}$ and another with negative weights $\mathscr{N} = \{p|w_p < 0\}$. The orbits in $\mathscr{P}$ and $\mathscr{N}$ cover approximately the same regions of state space

$$\sum_{p\in\mathscr{P}} w_p f_p(\mathbf{x}) \approx \sum_{n\in\mathscr{N}} w_n f_n(\mathbf{x}). \tag{38}$$

When either $\mathscr{N}$ or $\mathscr{P}$ have only one member, this sole orbit must be a long orbit that shadows all short orbits in the other set. Hence, marginal modes can shed light on how orbits shadow each other within a chaotic system. Given that UPOs have been shown to shadow each other also in fluid flows,[28] we expect such marginal modes to exist in applications to higher-dimensional systems as well.

When weights are both positive ($w_p \geq 0$, for all $p$) and normalized ($\sum_p w_p = 1$), each $w_p$ is interpretable as the probability that a randomly sampled chaotic state is well-approximated by a state from orbit $p$, and Eq. (3) is interpretable as a mixture model. This makes positivity and normalization enticing properties for weights to satisfy. Normalization is required of any collection of weights that

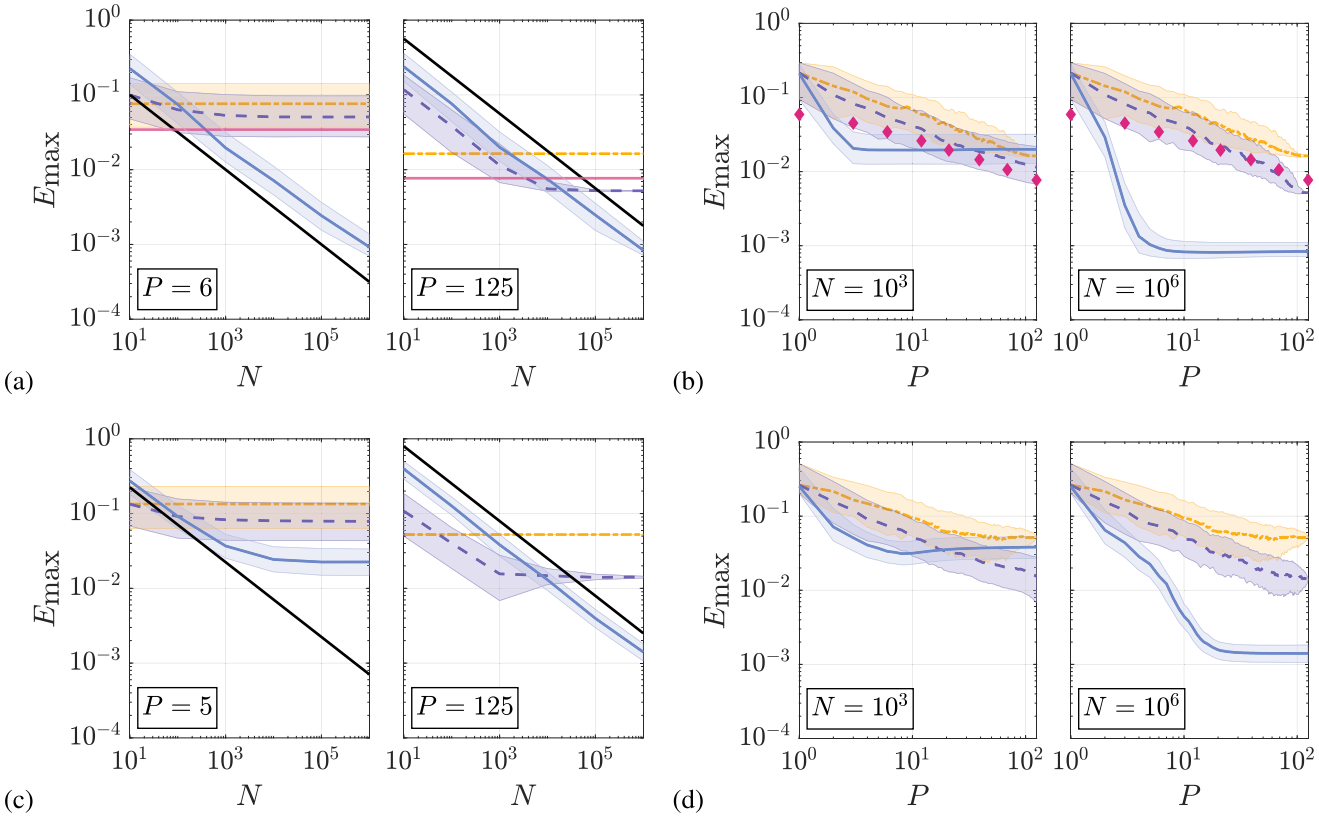


**FIG. 3.** The median (line) and interquartile range (shaded region) of the error $E_{\max}$ over $s$ and $r$ is plotted for orbits [panels (a) and (b)] and snippets [panels (c) and (d)]. In (a) and (c), $P$ is fixed and $N$ is varied. In (b) and (d), $N$ is fixed and $P$ is varied. Line style and color correspondence is as follows: LSW (solid, blue), Markov (dashed, purple), and uniform (dot-dashed, yellow). The error of POT for complete libraries is overlaid as magenta diamonds. The scaling $E_{\max} \sim N^{-1/2}$ is plotted as a solid black line.

accurately predict the time average of the constant observable. When the sets $\{S_p\}$ are non-overlapping, it follows that any weights satisfying Eq. (3) must also be positive. The SRB measure is a probability measure, meaning that $\mathbb{E}[I(\mathbf{x}, A)] \geq 0$ for any set $A \subset \mathscr{M}$. Applying Eq. (3) to the indicator function $I(\mathbf{x}, S_q)$, it follows that

$$\sum_{p=1}^{P} w_p \mathbb{E}_p[I(\mathbf{x}, S_q)] = w_q = \mathbb{E}[I(\mathbf{x}, S_q)] \geq 0 \qquad q = 1, \ldots, P.$$

However, if the sets $\{S_p\}$ are overlapping, then the optimal weights can be negative, even when both $\mu$ and all $\mu_p$ are probability measures. As a simple example, consider expanding the uniform probability measure on a set of four points $\mu = \{1/4, 1/4, 1/4, 1/4\}$ in terms of probability measures $\mu_1 = \{1/3, 1/3, 1/3, 0\}$, $\mu_2 = \{0, 1/3, 1/3, 1/3\}$, and $\mu_3 = \{0, 1/2, 1/2, 0\}$. It follows that

$$\mu = \frac{3}{4}\mu_1 + \frac{3}{4}\mu_2 - \frac{1}{2}\mu_3,$$

where the weight associated with $\mu_3$ is negative.

Markov weights are normalized and positive by definition. POT weights are guaranteed to be normalized; however, it is unproven whether or not POT weights are positive. Empirically, we find that POT weights are positive when computed on complete libraries, but can produce negative weights when evaluated on incomplete libraries. The Tikhonov-regularized solution used in Sec. III does not enforce positivity or normalization and is not guaranteed to produce weights that can be interpreted as probabilities. As mentioned, we find that the solution found by Tikhonov regularization happens to be normalized to nearly four significant digits. Moreover, we observe that the Tikhonov-regularized solution becomes increasingly positive (and remains approximately normalized) as the regularization parameter $\alpha$ is increased. However, the residual of the original least squares problem, $\|A\vec{w} - \vec{b}\|$, increases with increasing $\alpha$. The weights resulting from Tikhonov regularization with $\alpha = 10^{-10}$ and $P = 125$ are plotted as a function of orbit stability in Fig. 4(a).

Is it possible to constrain the LSW solution to be both positive and normalized. That is, we may seek minima of the function

$$V = \|A\vec{w} - \vec{b}\|_2^2 \qquad \text{subject to} \quad w_p \geq 0, \quad \sum_{p=1}^{P} w_p = 1. \tag{39}$$

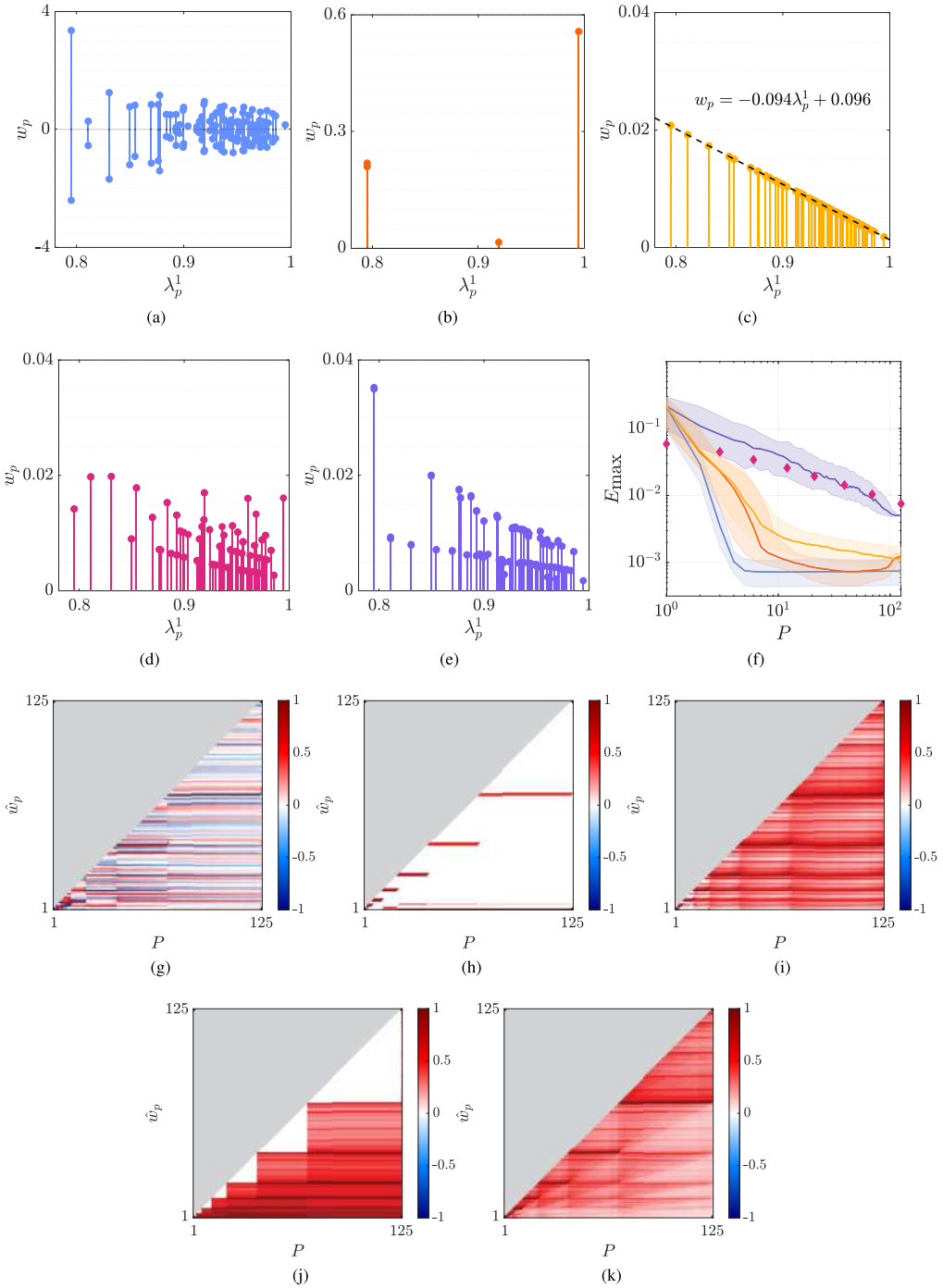


**FIG. 4.** A comparison of periodic orbits' weights for different methods. The (a) unconstrained, (b) sparse, and (c) positive-normalized least square weighting solutions. POT and Markov weights are shown for comparison in (d) and (e), respectively. In each panel (a)–(e), the weights $\{w_p\}_{p=1}^{125}$ are plotted as a function of the orbit's leading Floquet exponent $\lambda_p^1$. In (c), the dashed line shows a line of best fit. The error associated with each method is plotted as a function of $P$ in (f), in the same style as Fig. 3, with lines colored-coded to match the color of panels (a)–(e). Panels (g)–(k) show the redistribution of weight as $P$ increases, as described in the text. For visualization purposes, the weights are normalized $\hat{w}_p = w_p / \max_{q=1}^{P} |w_q|$ at every fixed value of $P$.

We do this two ways. First, we approximate a solution to Eq. (39) using MATLAB's built-in `lsqnonneg` function and normalizing the output. The output of `lsqnonneg` happens to be normalized to about four decimals places for all $P$, and the normalization we apply afterwards does not greatly perturb the solution away from a minimum of $V$. More importantly, as a consequence of MATLAB's implementation of `lsqnonneg`, these solutions are sparse, see Fig. 4(b). We found this to be the most reliable way to find a sparse solution.

We also use MATLAB's built-in `fmincon` function to minimize $V$ directly. The solution found by `fmincon` does depend on initial conditions. In particular, the weights resulting from minimization using the initial condition $w_p = 1/P$ are plotted in Fig. 4(c).

It has been posited before that $w_p$ should be inversely proportional to an orbit's stability;[30,47] intuitively, the more unstable an orbit is compared to others, the less frequently it should be visited by the chaotic trajectory, and the shorter the duration of each visit. Interestingly, almost all weighting schemes (save the sparse one) indicate that $|w_p|$ tends to decrease with an orbit's leading Floquet exponent $\lambda_p^1$, although the dependence is generally nonmonotonic. Surprisingly, the solution found by `fmincon` yields a simple linear relationship between $w_p$ and $\lambda_p^1$ (rather than $1/\lambda_p^1$, as has been suggested previously). Note that such linear relationship is only found for complete libraries (e.g., $P = 125$).

To illustrate how weight gets redistributed among orbits as orbits get added to the expansion, Fig. 4 also shows the distribution of $\{w_p\}_{p=1}^P$ as a function of increasing $P$ (over library $\mathscr{L}_1$ and $N = 10^6$). Panels (g)–(k) illustrate the unconstrained LSW, sparse LSW, normalized and positive LSW, POT, and Markov weightings, respectively. In each panel, color denotes the magnitude of $w_p$ for all orbits $p \leq P$ included in the expansion, with red (blue) indicating the positive (negative) weight. Orbits $p > P$ excluded at a given value of $P$ are grayed out. These plots illustrate that the distributions of weight, across all weighting methods, vary rather smoothly as orbits get added to the expansion, with notable discontinuities at values of $P$ corresponding to complete libraries of different symbolic lengths. The sparse LSW retains very particular orbits. Using the standard labeling convention for symbolic dynamics,[44] the sparse solution prefers to retain orbit $AB$—the shortest orbit of the Lorenz dynamics—along with highly asymmetric orbits of the form $A^nB$ and $B^nA$ for integer $n$. Since POT can only be evaluated for complete libraries, for every fixed value of $P$, Fig. 4(j) shows the weights for the largest complete library of size less than or equal to $P$.

Perhaps the most surprising result of this study is that Eq. (3) admits multiple, qualitatively distinct distributions of weights that yield accurate estimates of time averages, even when $P$ is small [see Figs. 4(a)–4(c) and 4(f)]. To the authors' knowledge, quantitative evidence of this has not been presented in the literature thus far. Most impressively, the sparse solution found by `lsqnonneg` achieves comparable accuracy to the unconstrained LSW solution and returns weights that may be interpreted as probabilities, while never requiring more than four orbits in the expansion.

## V. METHOD COMPARISON

There are currently three leading approaches to computing solutions to Eq. (3), including the Markov weighting, POT, a non-linear optimization approach based upon minimizing a Kullback–Leibler (KL) divergence[48]—which we did not compare to here—that LSW can be compared with. It is important to understand (i) the observables each of these methods aims to estimate and (ii) each method's computational cost.

Indeed, since Eq. (3) represents a *reduced order model* of $\mathbb{E}[\,\cdot\,]$—one with only $P$ degrees of freedom—$\hat{\mathbb{E}}[\,\cdot\,]$ cannot possibly predict every observable equally well. Each choice of weights biases the model toward predicting certain observables *exactly*. Equation (29) makes explicit the class of smooth observables targeted by the LSW. Moreover, Table I indicates that targeting observables in $\mathscr{A}$ is a good prior for extrapolating the time averages over other observables well. Figure 4(f) indicates that LSW continues to extrapolate the time averages of observables outside $\mathscr{A}$ even when the weights are made sparse.

The Markov weighting targets piecewise constant functions of the form

$$a(\mathbf{x}) = \sum_{p=1}^{P} c_p I(\mathbf{x}, S_p), \tag{40}$$

for arbitrary coefficients $c_p$. To see this, notice that when $\{S_p\}_{p=1}^P$ form a Markov partition, one can derive the Markov weights by enforcing that Eq. (3) perfectly predicts the averages of each $I(\mathbf{x}, S_p)$,

$$\mathbb{E}[I(\mathbf{x}, S_q)] = \sum_{p=1}^{P} w_p \mathbb{E}_p[I(\mathbf{x}, S_q)] = w_q, \tag{41}$$

since $\mathbb{E}_p[I(\mathbf{x}, S_q)] = \delta_{pq}$. As a direct consequence, it follows for any piecewise-constant $a(\mathbf{x})$ that

$$\mathbb{E}[a] = \sum_{p=1}^{P} c_p \mathbb{E}[I(\mathbf{x}, S_p)] = \sum_{p=1}^{P} c_p w_p = \sum_{p=1}^{P} w_p \mathbb{E}_p[a]. \tag{42}$$

Physical observables—such as the kinetic energy or dissipation rate of fluid flows—are often continuous functions of $\mathbf{x}$ that lie outside the space of piecewise-constant functions targeted by Markov weights. However, as $P \to \infty$, these piecewise-constant functions can approximate such observables increasingly well.

Although error bounds for predicting time averages of arbitrary smooth observables using POT have been established,[12,49] it is to the authors' knowledge unknown which space of observables is predicted exactly by a $P$-orbit POT expansion. At minimum, when the library is complete, POT will predict with zero error time averages that vanish as a consequence of symmetries in the governing equations (cf. Table I).

Redfern *et al.*[48] compute weights based on the probability density function (PDF) of a single observable (such as energy dissipation rate $D$ for a fluid flow). A KL divergence-based approach aims to match the PDF $\Gamma(D)$ computed over a chaotic trajectory to PDFs $\Gamma_p(D)$ computed over $P$ periodic orbits. If $\Gamma(D)$ is estimated by $\sum_{p=1}^P w_p \Gamma_p(D)$ sufficiently well, one can evaluate an accurate average of *any* function of the observable $D$. However, this approach has difficulty in extrapolating to averages of observables that cannot be expressed as functions of $D$ (such as energy $E$ or energy input $I$ for fluid flows, cf. Table 1 in Ref. 48). Moreover, this extrapolation error grows drastically as $P$ decreases. By biasing all $P$ degrees

of freedom in the reduced order model to capture the fine details of $\Gamma(D)$, information about other observables is forgone. To combat this, weights could be trained on multivariate PDFs, although this is computationally costly, as discussed below.

Indeed, one must also consider the effort needed to compute the weights. Suppose orbits (or, in the case of LSW, snippets) are sampled discretely using, on average, $n$ snapshots in time and the chaotic trajectory is sampled using $N$ snapshots in time. In $d$ dimensions, the computational cost of evaluating each $A_{pq}$ is then $O(dn^2)$. Similarly, the computational cost associated with evaluating each $b_p$ is $O(dnN)$, and the computational cost of solving the linear system is $O(P^3)$. Thus, the total computational cost of the LSW method is $O(dn(nP^2 + NP) + P^3)$. In the $N \gg P$ limit considered here, the cost scales as $O(dnNP)$.

This cost may be compared to that of Markov partition-based methods which scale like $O(dNP)$ and are cheaper by a factor of $O(n)$.

The cost of the KL divergence-based method is hard to estimate precisely since it is iterative. Assuming the cost of evaluating $a(\mathbf{x})$ is $O(d)$, the cost of generating a single histogram over $B$ bins is $O(dN)$, where $N \gg B$ such that the histogram is well-resolved. If the method were trained on $k$-dimensional histograms, in order to ameliorate the extrapolation error, one requires $N \gg B^k$. Each iteration of gradient descent will cost $O(B^kP)$. For a gradient descent scheme requiring $I$ iterations to converge, the total cost scales like $O(dN + IB^kP)$. For low $k$, the cost $O(dN)$ is cheaper than Markov-based methods by a factor $O(P)$. For high $k$, the cost $O(IB^kP)$ becomes computationally prohibitive.

POT is also iterative, since it requires using a root-finding scheme to determine the escape rate of the $P$ orbits (see Appendix A). A benefit of POT is that, in most Axiom A systems, it is known *a priori* which orbits are embedded in the chaotic set. As a result, the weights (and from them averages) can be computed without ever computing a chaotic trajectory. Assuming a complete library of UPOs with known stability spectra, the cost is just $O(IP \log P)$, where $I$ is the number of iterations required by the root-finding method. In this idealized setting, POT does not scale with $N$, providing it the lowest computational cost by a large margin. However, in practice, this conclusion is highly misleading. Generally, it is very difficult to determine which periodic orbits are embedded in the chaotic set; for instance, periodic orbits that are embedded in the chaotic set at one value of parameters may not be at others.[50] Hence, in non-Axiom-A systems, POT also requires computing a chaotic trajectory to identify the relevant periodic orbits, e.g., through shadowing analysis,[26] making the cost of POT scale with $N$ just like the other empirical approaches. The exact complexity of the method will depend on how the dynamical relevance of orbits is verified.

In summary, although more computationally expensive than its alternatives, LSW generally yields estimates of time averages that are at least an order of magnitude more accurate and yields a reduced order model of $\mathbb{E}$ that extrapolates more accurately to non-targeted observables. It is worth appreciating that the LSW estimator of $\mathbb{E}$ appears to describe temporal averages over a broad class of observables with exceptional accuracy, even from a small number of UPOs of snippets; in contrast to other methods, its accuracy is largely insensitive to library sparsity, as illustrated by Figs. 3(b) and 3(d) for $P \gtrsim 13$.

## VI. CONCLUSION

To summarize, we have derived a general method for performing least squares regression on linear relations of measures and used it to construct an accurate approximation of infinite-time averages for dynamical systems in terms of a weighted average over as few as four periodic orbits or ten snippets of chaotic trajectory. Thus, the approach introduced here works almost equally well whether or not any periodic orbits are known.

The weights are given by a least squares solution to a linear system of equations, and, unlike those predicted by period orbit theory, are quite straightforward to compute. As demonstrated here, enforcing constraints on the weights, such as positivity and normalization, is also simple using off-the-shelf linear programming algorithms.

We have also identified the subspace of observables over which the least squares weighting yields exact predictions. Even outside of this subspace, the accuracy of the method tends to be an order of magnitude better than state-of-the-art methods, and more quickly convergent in the number of orbits or snippets used. When applied to periodic orbits, the method can accurately estimate Lyapunov exponents and Kaplan–Yorke dimension of the chaotic set. Surprisingly, the method discovers *multiple* weightings that outperform state-of-the-art methods.

Given that the least squares weighting is based on Kriging, it is expected to scale quite well to high- and even infinite-dimensional state spaces, such as those needed to describe fluid, plasma, or optical turbulence, although this has yet to be verified. In such applications, few if any periodic orbits are typically known, so the ability of our approach to work when applied to few orbits and snippets of chaotic trajectory makes it uniquely appealing.

## ACKNOWLEDGMENTS

We gratefully acknowledge financial support from the National Science Foundation under Grant No. 2032657.

## AUTHOR DECLARATIONS

### Conflict of Interest

The authors have no conflicts to disclose.

### Author Contributions

**Joshua L. Pughe-Sanford:** Conceptualization (lead); Formal analysis (equal); Software (lead); Writing – original draft (equal); Writing – review & editing (equal). **Sam Quinn:** Conceptualization (supporting); Formal analysis (equal); Software (supporting); Writing – original draft (equal); Writing – review & editing (equal). **Teodor Balabanski:** Formal analysis (supporting); Software (supporting). **Roman O. Grigoriev:** Supervision (lead); Writing – original draft (equal); Writing – review & editing (equal).

## DATA AVAILABILITY

The code and data that support the findings of this study are openly available in Github, Ref. 51. The algorithm used for computing POT weights is described in Appendix A.

## APPENDIX A: COMPUTING POT WEIGHTS NUMERICALLY

The theoretical foundations of POT are summarized by Cvitanović,[16] and Chaosbook.org[12] provides a detailed reference text on the subject. Eckhardt and Ott[14] applied POT to the Lorenz 1963 system. Here, we provide a short summary of the essential steps. The following algorithm works best for systems with symbolic dynamics; however, averages can still be computed from POT without one as discussed in Ch. 23.7 of Ref. 12. Viswanath[44] provides a good (and visual) description of symbolic dynamics in the context of Lorenz.

Given a library of periodic orbits that contains all orbits up to symbolic length $n$, the temporal average is approximated as

$$\mathbb{E}[a] \approx -\left.\frac{\partial_\beta F_n(s,\beta)}{\partial_s F_n(s,\beta)}\right|_{\substack{s=s_0,\\ \beta=0}}, \tag{A1}$$

where $F_n(s,\beta)$—which will be defined below—is a function that approximates a spectral determinant (see Ch. 22 of Ref. 12), and $s_0$ is the real zero of $F_n$, i.e., $F_n(s_0,0)=0$. Equality is reached in Eq. (A1) in the limit $n\to\infty$. While POT does not define the weights $w_p$ explicitly, they may be computed from Eq. (A1),

$$w_p = -\left.\frac{\partial_{\mathbb{E}_p[a]}\partial_\beta F_n(s,\beta)}{\partial_s F_n(s,\beta)}\right|_{\substack{s=s_0,\\ \beta=0}}. \tag{A2}$$

While this definition appears to depend on the choice of observable, this dependence vanishes at $\beta=0$ [cf. Eq. (A8)].

The spectral determinant can be computed using the trace coefficients (see Ch. 23.2.2 of Ref. 12),

$$C_j = -\sum_{n_p r=j}\frac{1}{r}\frac{e^{-rT_p\left(s-\beta\,\mathbb{E}_p[a]\right)}}{\left|\det\left(1-(M_p)^r\right)\right|}, \tag{A3}$$

where the sum is over all prime orbits $p$ such that $j=rn_p$, where $r=1,2,\ldots$ and $n_p$ is the symbol length of orbit $p$. Notice that $C_j$ sums over all periodic orbits of length $j$ *and* repeats of shorter orbits which, when repeated $r$ times, have symbol length $j$. $M_p$ is the transverse monodromy matrix (see Ch. 21.2.2 of Ref. 12) of orbit $p$, over one period. For the Lorenz system, $\det(1-(M_p)^r)\approx 1-e^{rT_p\lambda_p}$, where $\lambda_p$ and $T_p$ are the top Floquet exponent and period of orbit $p$, respectively.

The spectral determinant is then

$$F_n(s,\beta) = 1+\sum_{j=1}^{n} Q_j, \tag{A4}$$

with the help of a recurrence relation

$$Q_j = C_j + \sum_{i=1}^{j-1}\frac{j-i}{j}C_{j-i}Q_i, \tag{A5}$$

where $Q_1=C_1$. The partial derivatives of $F_n$ are computed via the chain rule with

$$\left.\partial_s C_j\right|_{\beta=0} = \sum_{n_p r=j}\frac{e^{-rT_p s}}{\left|1-e^{rT_p\lambda_p}\right|}T_p, \tag{A6}$$

$$\left.\partial_\beta C_j\right|_{\beta=0} = -\sum_{n_p r=j}\frac{e^{-rT_p s}}{\left|1-e^{rT_p\lambda_p}\right|}T_p\mathbb{E}_p[a], \tag{A7}$$

$$\left.\partial_{\mathbb{E}_p[a]}\partial_\beta C_j\right|_{\beta=0} = -\sum_{n_p r=j}\frac{e^{-rT_p s}}{\left|1-e^{rT_p\lambda_p}\right|}T_p, \tag{A8}$$

$$\left.\partial_{\mathbb{E}_p[a]} C_j\right|_{\beta=0} = 0. \tag{A9}$$

In Eq. (A8), note that $p$ is fixed and the sum runs only over $r$. The root $s_0$ is found using Newton's method by iterating the relation

$$s_0 = s_0 - \frac{F_n(s_0,0)}{\partial_s F_n(s_0,0)}, \tag{A10}$$

until $|F_n(s_0,0)|<\varepsilon$ using $s_0=0$ as the initial guess. We use $\varepsilon=10^{-8}$.

## APPENDIX B: LYAPUNOV EXPONENTS AND KAPLAN–YORKE DIMENSION

The Lyapunov exponents $\lambda^1=0.905\,66(7)$, $\lambda^2=0$, and $\lambda^3=-14.572\,33(7)$ of the Lorenz 1963 system were computed by Viswanath.[52] These exponents are defined as the long-time limits of the logarithms of the singular values of the Jacobian matrix $J(t)=\partial\mathbf{x}(t)/\partial\mathbf{x}(0)$. Their values are independent of the initial condition $\mathbf{x}(0)$ for almost all $\mathbf{x}(0)\in\mathscr{X}$, but may differ when $\mathbf{x}(0)$ lies on a periodic orbit. The values reported above are those computed over $\mathscr{X}$.

Computing the singular values over an infinitely long trajectory is numerically challenging. However, it has been argued that the real parts of the logarithms of the eigenvalues of $J(t)$ converge to the Lyapunov exponents in the limit $t\to\infty$.[53] When $\mathbf{x}(0)$ lies on a periodic orbit, these eigenvalue-based exponents—known as Floquet exponents—are especially tractable, since they can be computed from the finite-time monodromy matrix $J(T_p)$.

This motivates estimating the $i$th Lyapunov exponent by a weighted average Floquet exponent,

$$\hat{\lambda}^i \equiv \sum_{p=1}^{P} w_p\lambda_p^i \approx \lambda^i, \tag{B1}$$

where $\lambda_p^i$ is the real part of the $i$th Floquet exponent for orbit $p$. Indeed, for Axiom A systems, this expansion is known to converge in the limit $P\to\infty$.[12,14] For the Lorenz model, this relation is also trivially true for $\lambda^2$, since all $\lambda_p^2$ are also zero as a consequence of time-translation invariance.

The Kaplan–Yorke dimension of a chaotic set is defined as the interpolated zero of the function $\zeta(d)=\sum_{i=1}^{d}\lambda^i$[17,54] and describes the dimension of a subset $S\subset\mathscr{M}$ whose volume is preserved under evolution. In the Lorenz model, this interpolated zero occurs at

$$d_{KY} = 2+\frac{\lambda^1+\lambda^2}{|\lambda^3|} = 2.062\,149(5), \tag{B2}$$

where the numeric value of $d_{KY}$ (and its error) is propagated from the values of $\lambda^i$ reported above. This dimension can be estimated by

$$\hat{d}_{KY} = 2 + \frac{\hat{\lambda}^1 + \hat{\lambda}^2}{|\hat{\lambda}^3|}. \tag{B3}$$

## REFERENCES


[1]A. Christensen, K. Tsiaras, J. Murawski, Y. Hatzonikolakis, J. She, M. St. John, U. Lips, and R. Brouwer, "A cross disciplinary framework for cost-benefit optimization of marine litter cleanup at regional scale," Front. Mar. Sci. **8**, 744208 (2021).
[2]T. Yamada, S.-I. Itoh, T. Maruta, N. Kasuya, Y. Nagashima, S. Shinohara, K. Terasaka, M. Yagi, S. Inagaki, Y. Kawai, A. Fujisawa, and K. Itoh, "Anatomy of plasma turbulence," Nat. Phys. **4**, 721–725 (2008).
[3]A. Picozzi, J. Garnier, T. Hansson, P. Suret, S. Randoux, G. Millot, and D. Christodoulides, "Optical wave turbulence: Towards a unified nonequilibrium thermodynamic formulation of statistical nonlinear optics," Phys. Rep. **542**, 1–132 (2014).
[4]E. N. Lorenz, "Deterministic nonperiodic flow," J. Atmos. Sci. **20**, 130–141 (1963).
[5]G. D. Birkhoff, "Proof of the ergodic theorem," Proc. Natl. Acad. Sci. USA **17**, 656–660 (1931).
[6]Y. G. Sinai, "Gibbs measures in ergodic theory," Russ. Math. Surv. **27**, 21–69 (1972).
[7]D. Ruelle, "A measure associated with axiom-a attractors," Am. J. Math. **98**, 619 (1976).
[8]R. E. Bowen, J.-R. Chazottes, and D. Ruelle, "Ergodic theory of axiom a diffeomorphisms," in *Equilibrium States and the Ergodic Theory of Anosov Diffeomorphisms* (Springer, Berlin, 2008), Vol. 470, pp. 61–73.
[9]R. O. Grigoriev and H. G. Schuster, "Solvable model for spatiotemporal chaos," Phys. Rev. E **57**, 388 (1998).
[10]J. Graham, K. Kanov, X. I. A. Yang, M. Lee, N. Malaya, C. C. Lalescu, R. Burns, G. Eyink, A. Szalay, R. D. Moser, and C. Meneveau, "A Web services accessible database of turbulent channel flow and its use for testing a new integral wall model for LES," J. Turbul. **17**, 181–215 (2016).
[11]S. Ulam, *Problems in Modern Mathematics* (Courier Corporation, 2004).
[12]P. Cvitanović, R. Artuso, R. Mainieri, G. Tanner, and G. Vattay, *Chaos: Classical and Quantum* (Niels Bohr Institute, 2020).
[13]D. Fernex, B. R. Noack, and R. Semaan, "Cluster-based network modeling—From snapshots to complex dynamical systems," Sci. Adv. **7**, eabf5006 (2021).
[14]B. Eckhardt and G. Ott, "Periodic orbit analysis of the Lorenz attractor," Z. Phys. B **93**, 259–266 (1994).
[15]P. Cvitanović, "Dynamical averaging in terms of periodic orbits," Physica D **83**, 109–123 (1995).
[16]P. Cvitanović, "Trace formulas in classical dynamical systems," in *Supersymmetry and Trace Formulae*, edited by I. V. Lerner, J. P. Keating, and D. E. Khmelnitskii (Springer US, 1999), Vol. 370, pp. 85–102.
[17]J. P. Eckmann and D. Ruelle, "Ergodic theory of chaos and strange attractors," Rev. Mod. Phys. **57**, 617–656 (1985).
[18]S. Smale, "Differentiable dynamical systems," Bull. Am. Math. Soc. **73**, 747–817 (1967).
[19]P. Cvitanović, R. Artuso, R. Mainieri, G. Tanner, and G. Vattay, *Chaos: Classical and Quantum* (Niels Bohr Institute, 2020), Section 23.5.
[20]N. B. Budanur, D. Borrero-Echeverry, and P. Cvitanović, "Periodic orbit analysis of a system with continuous symmetry—A tutorial," Chaos **25**, 073112 (2015).
[21]F. Waleffe, "Exact coherent structures in channel flow," J. Fluid Mech. **435**, 93–102 (2001).
[22]S. Toh and T. Itano, "A periodic-like solution in channel flow," J. Fluid Mech. **481**, 67–76 (2003).
[23]M. Chantry, A. P. Willis, and R. R. Kerswell, "Genesis of streamwise-localized solutions from globally periodic traveling waves in pipe flow," Phys. Rev. Lett. **112**, 164501 (2014).
[24]A. Shekar and M. D. Graham, "Exact coherent states with hairpin-like vortex structure in channel flow," J. Fluid Mech. **849**, 76–89 (2018).
[25]N. B. Budanur, K. Y. Short, M. Farazmand, A. P. Willis, and P. Cvitanović, "Relative periodic orbits form the backbone of turbulent pipe flow," J. Fluid Mech. **833**, 274–301 (2017).
[26]M. C. Krygier, J. L. Pughe-Sanford, and R. O. Grigoriev, "Exact coherent structures and shadowing in turbulent Taylor–Couette flow," J. Fluid Mech. **923**, A7 (2021).
[27]C. J. Crowley, J. L. Pughe-Sanford, W. Toler, M. C. Krygier, R. O. Grigoriev, and M. F. Schatz, "Turbulence tracks recurrent solutions," Proc. Natl. Acad. Sci. U.S.A. **119**, e2120665119 (2022).
[28]C. J. Crowley, J. L. Pughe-Sanford, W. Toler, R. O. Grigoriev, and M. F. Schatz, "Observing a dynamical skeleton of turbulence in Taylor–Couette flow experiments," Philos. Trans. R. Soc. A **381**, 20220137 (2023).
[29]G. Kawahara and S. Kida, "Periodic motion embedded in plane couette turbulence: Regeneration cycle and burst," J. Fluid Mech. **449**, 291–300 (2001).
[30]G. J. Chandler and R. R. Kerswell, "Invariant recurrent solutions embedded in a turbulent two-dimensional Kolmogorov flow," J. Fluid Mech. **722**, 554–595 (2013).
[31]B. Suri, L. Kageorge, R. O. Grigoriev, and M. F. Schatz, "Capturing turbulent dynamics and statistics in experiments with unstable periodic orbits," Phys. Rev. Lett. **125**, 064501 (2020).
[32]J. Page, P. Norgaard, M. P. Brenner, and R. R. Kerswell, "Recurrent flow patterns as a basis for two-dimensional turbulence: Predicting statistics from structures," Proc. Natl. Acad. Sci. U.S.A. **121**, e2320007121 (2024).
[33]G. Yalnız, B. Hof, and N. B. Budanur, "Coarse graining the state space of a turbulent flow using periodic orbits," Phys. Rev. Lett. **126**, 244502 (2021).
[34]J. F. Gibson, J. Halcrow, and P. Cvitanović, "Equilibrium and travelling-wave solutions of plane Couette flow," J. Fluid Mech. **638**, 243–266 (2009).
[35]J. S. Park and M. D. Graham, "Exact coherent states and connections to turbulent dynamics in minimal channel flow," J. Fluid Mech. **782**, 430–454 (2015).
[36]D. A. Lind and B. Marcus, *An Introduction to Symbolic Dynamics and Coding* (Cambridge University Press, USA, 1995).
[37]P. Cvitanović, R. Artuso, R. Mainieri, G. Tanner, and G. Vattay, *Chaos: Classical and Quantum* (Niels Bohr Institute, 2020), Section 23.1.
[38]C. C. Maiocchi, V. Lucarini, and A. Gritsun, "Decomposing the dynamics of the Lorenz 1963 model using unstable periodic orbits: Averages, transitions, and quasi-invariant sets," Chaos **32**, 033129 (2022).
[39]N. Cressie, "The origins of Kriging," Math. Geol. **22**, 239–252 (1990).
[40]T. Wang, Z. Chen, G. Li, J. He, C. Liu, and X. Du, "A novel method for high-dimensional reliability analysis based on activity score and adaptive Kriging," Reliab. Eng. Syst. Saf. **241**, 109643 (2024).
[41]H. Nan, H. Liang, H. Di, and H. Li, "A gradient-assisted learning strategy of Kriging model for robust design optimization," Reliab. Eng. Syst. Saf. **244**, 109944 (2024).
[42]Z. Song, Z. Liu, H. Zhang, and P. Zhu, "An improved sufficient dimension reduction-based Kriging modeling method for high-dimensional evaluation-expensive problems," Comput. Methods Appl. Mech. Eng. **418**, 116544 (2024).
[43]J. M. Heninger, D. Lippolis, and P. Cvitanović, "Neighborhoods of periodic orbits and the stationary distribution of a noisy chaotic system," Phys. Rev. E **92**, 062922 (2015).
[44]D. Viswanath, "Symbolic dynamics and periodic orbits of the Lorenz attractor," Nonlinearity **16**, 1035–1056 (2003).
[45]Original orbit data are available at https://websites.umich.edu/~divakar/lorenz/index.html.
[46]D. Lippolis and P. Cvitanović, "How well can one resolve the state space of a chaotic map?" Phys. Rev. Lett. **104**, 014101 (2010).
[47]E. Kazantsev, "Unstable periodic orbits and attractor of the barotropic ocean model," Nonlinear Processes Geophys. **5**, 193–208 (1998).
[48]E. M. Redfern, A. L. Lazer, and D. Lucas, "Dynamically relevant recurrent flows obtained via a nonlinear recurrence function from two-dimensional turbulence," Phys. Rev. Fluids **9**, 124401 (2024).
[49]D. Ruelle, *Thermodynamic Formalism: The Mathematical Structure of Equilibrium Statistical Mechanics*, Cambridge Mathematical Library, 2nd ed. (Cambridge University Press, 2004).

[50] A. Cleary and J. Page, "Dynamical relevance of periodic orbits under increasing Reynolds number and connections to inviscid dynamics," arXiv:2502.06475 (2025).
[51] See https://github.com/cdggt/lorenz-lsw for all scripts necessary to fully reproduce the data and figures used in the manuscript.
[52] D. Viswanath, "Lyapunov exponents from random Fibonacci sequences to the Lorenz equations," Ph.D. thesis (Cornell University, 1998).
[53] I. Goldhirsch, P.-L. Sulem, and S. A. Orszag, "Stability and Lyapunov stability of dynamical systems: A differential approach and a numerical method," Physica D **27**, 311–337 (1987).
[54] J. L. Kaplan and J. A. Yorke, "Chaotic behavior of multidimensional difference equations," in *Functional Differential Equations and Approximation of Fixed Points*, edited by H.-O. Peitgen and H.-O. Walther (Springer, Berlin, 1979), pp. 204–227.