Stabilizing embedology: Geometry-preserving delay-coordinate maps

Armin Eftekhari, Han Lun Yap, Michael B. Wakin, and Christopher J. Rozell

Stabilizing Embedology: Geometry-Preserving Delay-Coordinate Maps

Armin Eftekhari,[1, *] Han Lun Yap,[2] Michael B. Wakin,[3] and Christopher J. Rozell[4, †]

[1]*Alan Turing Institute, British Library*

96 Euston Road, London NW1 2DB, UK

[2]*DSO National Laboratories of Singapore*

[3]*Department of Electrical Engineering and Computer Science*

Colorado School of Mines

1500 Illinois St., Golden, CO 80401, USA

[4]*School of Electrical and Computer Engineering*

Georgia Institute of Technology

777 Atlantic Dr. NW, Atlanta, GA 30332-0250, USA

(Dated: October 3, 2017)

Abstract

Delay-coordinate mapping is an effective and widely used technique for reconstructing and analyzing the dynamics of a nonlinear system based on time-series outputs. The efficacy of delay-coordinate mapping has long been supported by Takens' embedding theorem, which guarantees that delay-coordinate maps use the time-series output to provide a reconstruction of the hidden state space that is a one-to-one embedding of the system's attractor. While this topological guarantee ensures that distinct points in the reconstruction correspond to distinct points in the original state space, it does not characterize the quality of this embedding or illuminate how the specific parameters affect the reconstruction. In this paper, we extend Takens' result by establishing conditions under which delay-coordinate mapping is guaranteed to provide a stable embedding of a system's attractor. Beyond only preserving the attractor topology, a stable embedding preserves the attractor geometry by ensuring that distances between points in the state space are approximately preserved. In particular, we find that delay-coordinate mapping stably embeds an attractor of a dynamical system if the stable rank of the system is large enough to be proportional to the dimension of the attractor. The stable rank reflects the relation between the sampling interval and the number of delays in delay-coordinate mapping. Our theoretical findings give guidance to choosing system parameters, echoing the trade-off between irrelevancy and redundancy that has been heuristically investigated in the literature. Our initial result is stated for attractors that are smooth submanifolds of Euclidean space, with extensions provided for the case of strange attractors.

---

[*] The first two authors contributed equally, as did the last two authors.

[†] Corresponding author:crozell@gatech.edu

I. INTRODUCTION

A. Motivation and Contribution

Modern science is ingrained with the premise that repeated observations of a dynamic phenomenon can help us understand its underlying mechanisms and predict its future behavior. While this idea dates back to ancient times with the observation of sunspots [1], today we model the behavior of a wide variety of measured phenomena from the life, physical, and social sciences [2–11] as observations arising from complex dynamical systems. Understanding and predicting a time series is often approached by postulating a structured model for a hidden dynamical system that drives the data generation. Linear statistical models were used in early work [12] and are now reflected in standard tools such as the *autoregressive-moving-average* model and the *Kalman filter* (e.g., [13]). More recently, the field of *nonlinear time-series analysis* models time-series data as observations of the state of a (possibly high-dimensional) deterministic nonlinear dynamical system [11]. While the underlying dynamical system may exhibit chaotic behavior, it is often postulated as being governed by an attractor that is a low-dimensional geometric subset of the state space.

Due to the low-dimensional behavior in the underlying state space, it is reasonable to postulate that temporal dependencies in time-series observations can provide some insight into the structure of the hidden dynamical system. This leads to a fundamental question: *How much information about a hidden dynamical system is available in time-series measurements of the system state?* The seminal *Takens' embedding theorem* [14, 15] asserts that (under very general conditions) it is possible to use the time-series data to reconstruct a state-space that is a topologically-equivalent image of the attractor through a simple procedure known as the delay-coordinate map. Indeed, many algorithms for tasks such as time-series prediction and dimensionality estimation take inspiration and justification from this fundamental guarantee [16–24]. While the topological guarantee of Takens' theorem provides that the delay-coordinate map is one-to-one (i.e., distinct points in the reconstruction correspond to distinct points in the original state space), it does not speak to the overall quality of the reconstruction or how this quality is affected by specific details such as the algorithm parameters, the measurement function, or the system characteristics.

Many fields of data science also rely on capturing low-dimensional structure from high-

dimensional data, and recent advances have shown the value of guaranteeing *geometric stability* of an embedding as a measure of quality for subsequent inference. In such a stable embedding, the embedding approximately preserves the distance between any two points in the data set of interest, which has proven to be valuable for robustness to imperfections in many forms (i.e., noise, numerical imprecision, etc.). In computer science, the Johnson-Lindenstrauss lemma constructs stable embeddings for finite point clouds using random linear projections [25]. In compressive sensing [26, 27], the Restricted Isometry Property (RIP) condition captures the notion of a stable embedding for sparse signal families, ensuring that signal reconstruction from random linear measurements is robust to noise and stable with respect to model nonidealities [28]. For dimensionality reduction with signal families belonging to low-dimensional manifolds and more general sets, various types of stable embeddings have been constructed using adaptive nonlinear techniques such as ISOMAP [29], adaptive linear techniques [30, 31], and nonadaptive linear techniques that again employ randomness [32–35].

The main contribution of this paper is to extend the notion of Takens' embedding theorem to stable embeddings, providing insight into the conditions for when time-series data can (and cannot) be used to reconstruct a geometry-preserving image of the attractor. In addition to providing the formal foundations to justify the numerical algorithms based on delay-coordinate mapping, these results also give guidance to practitioners about how algorithm and observation design choices affect the overall quality of the representation. In particular, examination of our main theoretical findings gives guidance to choosing these system parameters, echoing the trade-off between irrelevancy and redundancy that has been heuristically investigated in the literature. For clarity and to gain as much insight as possible, our main result is first described for attractors that are smooth submanifolds of the Euclidean space (similar to Takens' original result) and then extended to the case of strange attractors. The remainder of the Introduction will provide a simplified version of the main result to give the flavor of the contribution from this paper, with the full technical results given in Sections III C (smooth manifolds) and III E (strange attractors). To streamline readability as much as possible, the proofs and additional technical details are contained in appendices for the interested reader.

B. Delay-Coordinate Maps and Takens' Embedding Theorem

We consider $x(\cdot)$ as the trajectory of a dynamical system in the state space $\mathbb{R}^N$ such that $x(t) \in \mathbb{R}^N$ for $t \in [0, \infty)$. While the system has continuous underlying dynamics, we observe this system at a regular sampling interval $T > 0$. Given this sampling interval, one may define the discrete dynamics in terms of the *flow* $\phi_T : \mathbb{R}^N \to \mathbb{R}^N$ such that $x(t + T) = \phi_T(x(t))$. In words, $\phi_T(\cdot)$ moves the system state into the future by T. We assume that during the times of interest the state trajectory is contained within a low-dimensional *attractor* [11] $\mathbb{A}$ such that $x(t) \in \mathbb{A} \subset \mathbb{R}^N$ for $t \geq 0$. The attractor $\mathbb{A}$ is assumed to be a bounded, boundary-less, and smooth submanifold of $\mathbb{R}^N$ with $\dim(\mathbb{A}) < N$. The flow operator restricted to this attractor is a diffeomorphism on $\mathbb{A}$ so that there exists a smooth inverse $\phi_T^{-1}(x(t)) = x(t - T)$.

In applications of interest we often cannot directly observe this system state but rather receive indirect measurements via a scalar measurement function $h : \mathbb{A} \to \mathbb{R}$. This function generates a single scalar measurement at a regular sampling interval $T > 0$, producing the resulting discrete time series $\{s_i\}_{i \in \mathbb{N}} = \{h(x(iT))\}_i$, where each $s_i \in \mathbb{R}$. The goal is to "reconstruct" the hidden state trajectory $x(\cdot)$ given only $\{s_i\}_i$. To approach this task, consider the *delay-coordinate map* $F_{h,T,M} : \mathbb{A} \to \mathbb{R}^M$, defined for an integer number of delays M through the relation

$$F_{h,T,M}\left(x(iT)\right) = \begin{bmatrix} s_i \\ s_{i-1} \\ \vdots \\ s_{i-M+1} \end{bmatrix} = \begin{bmatrix} h(x(iT)) \\ h\left(x((i-1)T)\right) \\ \vdots \\ h\left(x((i-M+1)T)\right) \end{bmatrix} = \begin{bmatrix} h(x(iT)) \\ h\left(\phi_T^{-1}(x(iT))\right) \\ \vdots \\ h\left(\phi_T^{-M+1}(x(iT))\right) \end{bmatrix}. \quad (1)$$

Note that the delay-coordinate map is simply formed at a given time by stacking the last M observed time-series values into a vector. Commonly, $\mathbb{R}^M$ is referred to as the *reconstruction space*.

Takens' embedding theorem [14, 15] asserts that it is indeed possible to reconstruct the state space from the time-series data. With this setup, Takens' result roughly states that if $M > 2\dim(\mathbb{A})$, then the delay-coordinate map $F_{h,T,M}(\cdot)$ resulting from almost every smooth measurement function $h(\cdot)$ embeds the attractor $\mathbb{A}$ into the reconstruction space $\mathbb{R}^M$ (i.e., the delay-coordinate map forms a diffeomorphism for $\mathbb{A}$). In other words, the topology of the attractor $\mathbb{A}$ is preserved in the reconstruction space $\mathbb{R}^M$ under the delay-coordinate map, and
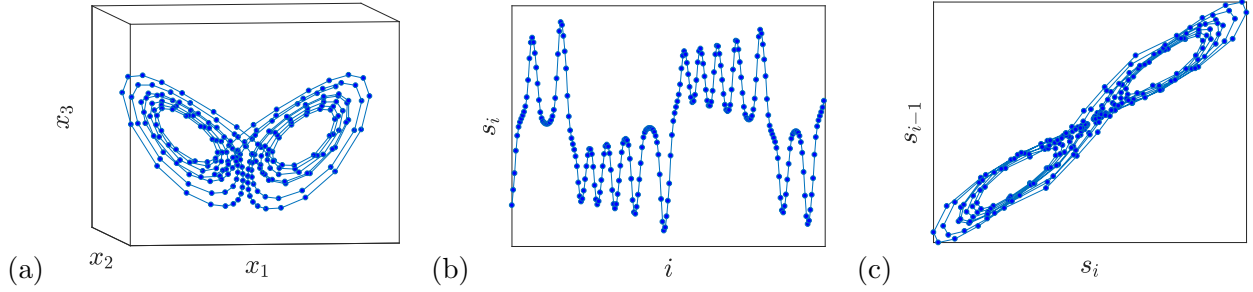
FIG. 1. *(a) The state space trajectory of the Lorenz attractor in $\mathbb{R}^3$, demonstrating the character-istic butterfly pattern. (b) The time series obtained by a measurement function that only keeps the x_1-coordinate of the trajectory. (c) The delay-coordinate map points with $M = 2$, recreating the butterfly pattern using only the time series.*

therefore the trajectory in the reconstruction space $F_{h,T,M}(x(\cdot))$ is (in principle) equivalent to the trajectory in the state space $x(\cdot)$. Figure 1 illustrates the concept of a delay-coordinate map in the case of the widely-known Lorenz attractor. Despite this embedding guarantee ensuring that no two points from the attractor map onto each other in the reconstruction, the mapping could be unstable in the sense that close points may map to points that are far away (and vice versa). While the simplified main result of this paper is presented for the case of attractors that are smooth submanifolds, the extensions presented in Section III E include strange attractors such as the Lorenz attractor.

C. Simplified Main Result

To quantify the quality of the embedding in the reconstruction space, we seek the stronger guarantee that the delay-coordinate map $F_{h,T,M}(\cdot)$ is a *stable embedding* of the attractor $\mathbb{A}$. By stable embedding, we mean that $F_{h,T,M}(\cdot)$ must act as a near-isometry on $\mathbb{A}$, in the sense that

$$\epsilon_l \leq \frac{\|F_{h,T,M}(x) - F_{h,T,M}(y)\|_2^2}{M\|x - y\|_2^2} \leq \epsilon_u, \qquad \forall x, y \in \mathbb{A}, \ x \neq y \tag{2}$$

for some *isometry constants* $0 < \epsilon_l \leq \epsilon_u < \infty$. Said another way, if $\epsilon_l \approx \epsilon_u$, the stable embedding condition of (2) guarantees that the delay-coordinate map preserves the *geometry* of the attractor (rather than merely its topology) by ensuring that pairwise distances between points on the attractor are approximately preserved in the reconstruction space. Since $x(\cdot) \subset \mathbb{A}$, the same would hold for the trajectory and its image, thereby guaranteeing the

quality of the trajectory embedding in the reconstruction space $\mathbb{R}^M$.

The main result of this paper is to determine the conditions on the attractor $\mathbb{A}$, measurement function $h(\cdot)$, number of delays M, and sampling interval T such that $F_{h,T,M}(\cdot)$ is a stable embedding of $\mathbb{A}$. This is a more ambitious objective than Takens' embedding theorem (leading naturally to more restrictive conditions), but with the benefit of quantifying the quality of the embedding and relating that quality to the problem-specific parameters. Roughly speaking, our main result shows that $F_{h,T,M}(\cdot)$ stably embeds $\mathbb{A}$ (in the sense of (2)) for most measurement functions h, provided that the following condition is satisfied:

$$
\mathrm{R}_{H,T,M}(\mathbb{A}) \gtrsim \dim(\mathbb{A}) \log \left(\frac{\mathrm{vol}(\mathbb{A})^{\frac{1}{\dim(\mathbb{A})}}}{\mathrm{rch}(\mathbb{A})} \right). \tag{3}
$$

Here, $\dim(\mathbb{A})$ and $\mathrm{vol}(\mathbb{A})$ are the dimension and volume of the attractor $\mathbb{A} \subset \mathbb{R}^N$, and $\mathrm{rch}(\mathbb{A})$ is an attribute of $\mathbb{A}$ that captures its geometric regularity. To quantify the notion of "most" measurement functions, our result is probabilistic and holds with high probability over measurement functions drawn from a rich probability model H. The *stable rank* $\mathrm{R}_{H,T,M}(\mathbb{A})$ of $\mathbb{A}$ quantifies the ability of the random measurement functions to observe the system attractor. Both reach and stable rank are well-studied concepts, and will be discussed (along with the detailed probability model H) in full technical detail later.

Typically, if a dynamical system is fairly "predictable", then $\mathrm{R}_{H,T,M}(\mathbb{A})$ grows proportionally with M as the number of delays grows. In this case, the delay-coordinate map stably embeds $\mathbb{A}$ when the number of delays scales linearly with the dimension of the attractor as in Takens' original theorem. On the other hand, if the dynamical system is highly unpredictable, then it is likely that $\mathrm{R}_{H,T,M}(\mathbb{A})$ plateaus rapidly with increasing M and it will be more difficult to stably embed this system through delay-coordinate mapping even with very long delay vectors. In Section III, the main contribution of this paper precisely quantifies these conditions governing the quality of the embedding from the delay-coordinate map. As we also discuss, these conditions have a natural interpretation in the context of classical empirical methods for choosing T and M.

7

II. BACKGROUND AND RELATED WORK

A. Takens' Embedding Theorem

To expound on the overview in Section I B, we turn our attention to a detailed technical statement of Takens' theorem [14] showing that the delay-coordinate map $F_{h,T,M}(\cdot)$ embeds the attractor $\mathbb{A}$ (and, of course, the trajectory $x(\cdot) \subset \mathbb{A}$).

Theorem II.1. (Takens' embedding theorem [14]) *Let $\mathbb{A} \subset \mathbb{R}^N$ be a smooth, bounded, and boundary-less submanifold of $\mathbb{R}^N$, and let $M > 2\dim(\mathbb{A})$ be an integer. For pairs (ϕ_T, h) where the flow $\phi_T : \mathbb{A} \to \mathbb{A}$ is a diffeomorphism on $\mathbb{A}$ and where $h : \mathbb{R}^N \to \mathbb{R}$ is a smooth measurement function, it is a generic property that the delay-coordinate map $F_{h,T,M}(\cdot)$ is an embedding (i.e., diffeomorphism) of $\mathbb{A} \subset \mathbb{R}^N$ into the reconstruction space $\mathbb{R}^M$.*

In this theorem, "generic" means that the set of pairs (ϕ_T, h) for which $F_{h,T,M}(\cdot)$ yields an embedding is open and dense in the set of all mappings. This topological notion of genericity was later extended to an "almost every" probabilistic argument by Sauer et al. [15]. In fact, the probe space framework developed in [15] was the inspiration for our analysis which involves drawing h randomly from a subspace of measurement functions (see Section III A). We also note that by relaxing the manifold assumption, Takens' theorem has also been generalized to cover embedding of *fractal sets* such as *strange attractors* [15] (formed by chaotic dynamical systems [11]) and embeddings of forced systems [36].

Note that, under Takens' theorem, $F_{h,T,M}(\mathbb{A}) \subset \mathbb{R}^M$ is diffeomorphic to $\mathbb{A} \subset \mathbb{R}^N$, so that the topology of the attractor $\mathbb{A}$ and the flow on this attractor are preserved under delay-coordinate mapping. In particular, we may trace $F_{h,T,M}(x(\cdot))$ with its samples and "reconstruct" the trajectory in the (often inaccessible) state space using only the time-series data. In fact, several important characterizations of dynamical systems are preserved under delay-coordinate mapping and can be computed directly in the reconstruction space, including the number/types of fixed points/orbits, the dimension of attractor (i.e., $\dim(F_{h,T,M}(\mathbb{A})) = \dim(\mathbb{A})$), and the Lyapunov exponents [11, 37]. Justified by these properties, the reconstruction space representation formed by the delay-coordinate map has been used for many practical time-series processing algorithms [38], including tasks such as prediction [11, 39], noise reduction [40], chaos synchronization and control [41–43], system identification [44], and detection of causality in complex networks [45].

8

While Takens' original theorem proves that the delay-coordinate map is theoretically equivalent to the attractor in the hidden state space, it may map close points far apart and far points close together. This warping, though topologically equivalent, means that even small changes in the reconstruction space representation (e.g., due to noise, etc.) can amount to arbitrarily large changes in the corresponding points in the state-space. One might ask: Are there any conditions where the delay-coordinate map is guaranteed to bound the errors due to noise? Will changes to the delay-coordinate map parameters (e.g., increasing the number of delays, decreasing the sampling interval), the system, or the measurement function affect the quality of the reconstruction or its tolerance to noise? Takens' original theorem does not address these issues, motivating our study of geometrically stable embeddings.

B. Related Work

Prior work by Casdagli et al. [16] begins to methodically address the issue of noise in delay embeddings by studying the effects of the sampling interval T and number of delays M on asymptotic quantities defined to capture the effects of noise on the delay-coordinate map. We note that, when the image of an attractor is warped or folded (and thus, not a stable embedding), noise sensitivity can be a problem as the conditional probability of the state given a noisy observation of the delay vector may be poorly localized. In addition, the conditional variance of $h(x((i+1)T))$ (the next value in the time series) may increase, which reduces the ability to predict the time series.

While some approaches have been developed to perform noise reduction in the reconstruction space [46], more generally, one finds a rich literature on methods of choosing the optimal T and/or M that account for noise by examining quantities that can typically be interpreted as having some dependence on the distortion of the attractor. To illustrate the concept (without claiming to be an exhaustive review), methods have been proposed to choose parameters by examining how they change the neighborhood relationships between points (e.g., the introduction of "false nearest neighbors") [47–49], geometric quantities (e.g., space filling) intended to separate trajectories [50, 51], test statistics proposed for determining whether the result is a valid embedding of the input [52], and statistics related to the predictive power of the time series (e.g., mutual information) [17, 19]. The work in [20] recommends the *mean orbital period* (approximated from the oscillations of the time se-

ries) as a reliable choice for the *window length* (i.e., TM), noting that most methods for choosing embedding parameters are based on empirical arguments, use arbitrary criteria, and ultimately do not guarantee good reconstructions. While not primarily introducing a method to choose parameters per se, our results are some of the first to provide a theoretical basis for proposed methods by explicitly illustrating the impact of parameter choices (and other problem-specific details) on a natural measure for assessing reconstruction quality: geometric stability.

Our approach to guaranteeing the stability of delay-coordinate mapping relies heavily on recent advances in the fields of compressive sensing and geometric functional analysis. As mentioned previously, a central condition in compressive sensing is the RIP, which requires a linear measurement operator to provide a stable embedding of the sparse signal family. Of particular interest in compressive sensing are randomized linear measurement operators. In particular, when the measurement operators are constructed randomly (e.g., as a random matrix populated with independent and identically distributed Gaussian entries), the RIP can be satisfied with high probability [53]. The basic compressed sensing results have been extended to various classes of structured randomized measurement functions [32, 54–56] as well as other low-dimensional models [57] such as smooth manifolds [33–35]. The present work is especially indebted to recent developments in geometric functional analysis which appeared first in [55] to establish the RIP for a class of structured random matrices. It is also worth noting the recent work [58], in which sparse recovery tools are used to help identify a dynamical system in spite of large erasures in the available data.

A study of stable delay-coordinate mapping for *linear* dynamical systems and with measurement functions that are *deterministic* and *linear* has previously appeared in [59]. The current result is a significant extension of this previous work by allowing general nonlinear systems and measurement functions that are both randomized and nonlinear. However, the main result in the present work has a similar flavor to [59], as both papers highlight cases where the embedding quality plateaus and cannot be improved by increasing the number of delays M.

C. Differential Geometry

Because we will consider attractors $\mathbb{A}$ that are submanifolds of $\mathbb{R}^N$, it is helpful to review the differential geometry concepts that characterize $\mathbb{A}$ and play a major role in the present results. The reader may also refer to [60] for a more comprehensive introduction.

To any point $x \in \mathbb{A}$ we can assign a *tangent subspace* $\mathbb{T}_x\mathbb{A} \subset \mathbb{R}^N$ comprised of the directions of all curves on $\mathbb{A}$ that are tangent to x. The linear subspace $\mathbb{T}_x\mathbb{A}$ has dimension $\dim(\mathbb{A})$ in $\mathbb{R}^N$, and the union of all tangent subspaces is called the *tangent bundle* of $\mathbb{A}$:

$$\mathbb{T}\mathbb{A} := \bigcup_{x \in \mathbb{A}} \{x\} \times \mathbb{T}_x\mathbb{A}.$$

Consider a smooth map $\psi : \mathbb{A} \to \mathbb{A}$. The derivative of this map at $x \in \mathbb{A}$ is the linear operator $D\psi(x) : \mathbb{T}_x \to \mathbb{T}_{\psi(x)}$ that satisfies

$$\lim_{\tau \to 0} \|\psi(x + \gamma(\tau)) - \psi(x) - [D\psi(x)](\gamma(\tau))\|_2 = 0, \tag{4}$$

for every smooth curve $\gamma : [-1, 1] \to \mathbb{A}$ that passes through x with $\gamma(0) = x$. The *normal subspace* $\mathbb{N}_x\mathbb{A}$ is the $(N - \dim(\mathbb{A}))$-dimensional orthogonal complement of $\mathbb{T}_x\mathbb{A}$ with respect to $\mathbb{R}^N$. The *normal bundle* of $\mathbb{A}$ is the union of all normal subspaces:

$$\mathbb{N}\mathbb{A} := \bigcup_{x \in \mathbb{A}} \{x\} \times \mathbb{N}_x\mathbb{A}.$$

For $r > 0$, we also let $\mathbb{N}^r\mathbb{A}$ denote the *open normal bundle* of $\mathbb{A}$ of radius r comprised of all normal vectors of length less than r. For example, when $\mathbb{A}$ is the unit circle in $\mathbb{R}^2$ and $r \in (0, 1)$, $\mathbb{N}^r\mathbb{A}$ may be identified with an annulus of width $2r$ (around the circle).

A *geodesic curve* on $\mathbb{A}$ is a smooth curve that minimizes the ℓ_2 distance between every pair of nearby points that it connects [60]. The *geodesic distance* between a pair of points on $\mathbb{A}$ is the length of the shortest geodesic curve that connects them. The ℓ_2 distance between points never exceeds their geodesic distance. Throughout, we assume that $\mathbb{A}$ is regular in the sense that, for some *geodesic regularity* $\mathrm{geo}(\mathbb{A}) \in [1, \infty)$, we have

$$\|x - y\|_2 \leq d_{\mathbb{A}}(x, y) \leq \mathrm{geo}(\mathbb{A})\|x - y\|_2, \qquad \forall x, y \in \mathbb{A}, \tag{5}$$

where $d_{\mathbb{A}}(x, y)$ stands for the geodesic distance between $x, y \in \mathbb{A}$. For a circle, $\mathrm{geo}(\mathbb{A}) = \frac{\pi}{2}$.

The *reach* measures the regularity of a manifold and is closely related to its *condition number* [33, 53, 61].

Definition II.1. (Reach of a manifold [62]) *Let $\mathbb{A}$ be a bounded and smooth submanifold of $\mathbb{R}^N$. The reach of $\mathbb{A}$ (denoted with $rch(\mathbb{A})$) is the largest number $r \geq 0$ having the following property: The open normal bundle about $\mathbb{A}$ of radius r is embedded in $\mathbb{R}^N$ for all $r < rch(\mathbb{A})$.*

In perhaps the simplest example, the reach of a circle with radius ρ is simply ρ. Reach controls both local and global properties of a manifold. Its role is summarized in two key relationships. First, the curvature of any unit-speed geodesic curve on $\mathbb{A}$ is bounded by $1/\text{rch}(\mathbb{A})$. Second, at long geodesic distances, reach controls how close the manifold may curve back upon itself. For example, supposing $x, y \in \mathbb{A}$ with $d_{\mathbb{A}}(x, y) > \text{rch}(\mathbb{A})$, it must hold that $\|x - y\|_2 > \text{rch}(\mathbb{A})/2$. See [33] for more details.

III. MAIN RESULT

We are now prepared to give a precise setup and statement of the result that was summarized in Section I C, along with additional interpretation and discussion.

A. Measurement Apparatus

We first set up our framework for choosing a measurement function $h(\cdot)$ that is used to observe the trajectory $x(\cdot)$. In general we seek a result in which the choice of measurement function is not specific and arbitrary measurement functions chosen according to some probability measure will work with overwhelming probability. To do this, inspired by an approach developed in [15], we limit the measurement function to some subset $\mathcal{H}$ of the space of all smooth functions. To establish this subset, we fix a finite collection of P *measurement basis functions* $h_p : \mathbb{A} \to \mathbb{R}$, $p \in \{1, 2, \ldots, P\}$. For any coefficient vector $\alpha \in \mathbb{R}^P$, we define a measurement function $h_\alpha : \mathbb{A} \to \mathbb{R}$ that is a corresponding linear combination of the measurement basis functions:

$$h_\alpha(\cdot) = \sum_{p=1}^{P} \alpha[p] h_p(\cdot). \tag{6}$$

We limit our attention to the class $\mathcal{H}$ of measurement functions formed by arbitrary linear combinations of this set of basis functions:

$$\mathcal{H} := \left\{ h_\alpha(\cdot) \, : \, \alpha \in \mathbb{R}^P \right\} = \text{span} \left[\{h_p(\cdot)\}_{p=1}^{P} \right]. \tag{7}$$

12

Note that while the sum in (6) is linear, each basis function can be nonlinear, resulting in a rich and flexible set of potential measurement functions. For two concrete examples,

- $\mathcal{H}$ is the class of all linear functions on $\mathbb{R}^N$ when $h_p(\cdot) = \langle \cdot, e_p \rangle$ for $p \in \{1, 2, \ldots, P\}$ with $P = N$. Here, $e_p \in \mathbb{R}^N$ is the p-th canonical vector in $\mathbb{R}^N$ (i.e., $e_p[p] = 1$ and $e_p[n] = 0$ when $n \neq p$).

- $\mathcal{H}$ is the set of all N-variate polynomials of degree K if $\{h_p(\cdot)\}_{p=1}^P$ is the set of all monomials of degree K with $P = \binom{K+N}{N}$.

Our main theorem will depend on certain properties of the measurement basis functions that are revealed by defining the map $H : \mathbb{A} \to \mathbb{R}^P$, where

$$H(x) := \begin{bmatrix} h_1(x) & h_2(x) & \cdots & h_P(x) \end{bmatrix}^* \in \mathbb{R}^P, \qquad \forall x \in \mathbb{A}. \tag{8}$$

The superscript * indicates the transpose of a matrix or vector. We will require that the measurement basis functions are sufficiently well behaved in that the following three assumptions on H are met:

A1. $H(\cdot)$ is a bi-Lipschitz map on $\mathbb{A}$, in the sense that $l_H \|x - y\|_2 \leq \|H(x) - H(y)\|_2 \leq u_H \|x - y\|_2$ for every pair $x, y \in \mathbb{A}$, and for some $l_H, u_H \in (0, \infty)$;

A2. $H(\cdot)$ is a diffeomorphism between $\mathbb{A}$ and $H(\mathbb{A})$, resulting in $H(\mathbb{A}) \subset \mathbb{R}^P$ being a bounded, boundary-less, and smooth submanifold of $\mathbb{R}^P$ with $\dim(H(\mathbb{A})) = \dim(\mathbb{A})$; and

A3. the nonzero singular values of $DH(x)$ belong to some interval $[\eta_{\min}, \eta_{\max}] \subset (0, \infty)$, where $DH(\cdot)$ is the derivative of H (see Section II C).

Under the above assumptions on the basis functions, the flow $\phi_T : \mathbb{A} \to \mathbb{A}$ in the state space naturally induces a flow $\phi_{H,T} : H(\mathbb{A}) \to H(\mathbb{A})$ (in $\mathbb{R}^P$) specified as

$$H(x) \overset{\phi_{H,T}(\cdot)}{\to} \phi_{H,T}(H(x)) := H(\phi_T(x)), \qquad \forall x \in \mathbb{A}. \tag{9}$$

As with the flow $\phi_T(\cdot)$, the induced flow $\phi_{H,T}(\cdot)$ is a diffeomorphism (but on $H(\mathbb{A})$ rather than $\mathbb{A}$).

Let $F_{h_\alpha, T, M}(\cdot)$ denote the delay-coordinate map formed with a measurement function $h_\alpha(\cdot) \in \mathcal{H}$, and let $F_{h_p, T, M}(\cdot)$ denote the delay-coordinate map associated with the p-th

basis function $h_p(\cdot)$. For $x \in \mathbb{A}$, we will find it useful to collect the components of the delay-coordinate map due to the different measurement basis functions and write them as columns of a matrix such that:

$$X_{H,T,M} := \begin{bmatrix} F_{h_1,T,M}(x) & F_{h_2,T,M}(x) & \cdots & F_{h_P,T,M}(x) \end{bmatrix} \in \mathbb{R}^{M \times P}. \tag{10}$$

Using (6), we can confirm the following useful identity:

$$F_{h_\alpha,T,M}(x) = X_{H,T,M}\alpha, \qquad \forall x \in \mathbb{A}, \ \alpha \in \mathbb{R}^P. \tag{11}$$

The introduction of $\phi_{H,T}(\cdot)$ above allows us to rewrite $X_{H,T,M} \in \mathbb{R}^{M \times P}$ (see (10)) as

$$\begin{aligned} X_{H,T,M} &= \begin{bmatrix} H(x) & H\left(\phi_T^{-1}(x)\right) & \cdots & H\left(\phi_T^{-M+1}(x)\right) \end{bmatrix}^* & \text{(see (8))} \\ &= \begin{bmatrix} H(x) & \phi_{H,T}^{-1}(H(x)) & \cdots & \phi_{H,T}^{-M+1}(H(x)) \end{bmatrix}^*. & \text{(see (9))} \end{aligned} \tag{12}$$

We also define the "trajectory attractor" as

$$\mathbb{A}_{H,T,M} := \left\{ \begin{bmatrix} H(x) \\ \phi_{H,T}^{-1}(H(x)) \\ \vdots \\ \phi_{H,T}^{-M+1}(H(x)) \end{bmatrix} : x \in \mathbb{A} \right\} \subset \mathbb{R}^{MP}. \tag{13}$$

Taken together, (11) and (12) show that the reconstruction vectors produced by the delay-coordinate map $F_{h_\alpha,T,M}(\cdot)$ can be viewed as linear operator (which depends on α) acting on points in the trajectory attractor. The reach of the trajectory attractor will play a role in our main result.

B. Stable Rank

Lastly, our main result depends on a certain quantity that summarizes the quality of the measurement apparatus for a given dynamical system. To that end, drawing from the scientific computing literature [63], we first define the *stable rank* of a matrix $A \in \mathbb{R}^{M \times P}$ as

$$R(A) := \frac{\|A\|_F^2}{\|A\|^2}, \tag{14}$$

where $\|A\|_F$ and $\|A\|$ are the Frobenius and spectral norms of A, respectively. It is straightforward to confirm that

$$1 \leq R(A) = \frac{\sum_i \sigma_i(A)^2}{\sigma_1(A)^2} \leq \text{rank}(A), \tag{15}$$

14

where $\sigma_1(A) \geq \sigma_2(A) \geq \cdots \geq 0$ are the singular values of A. In a sense, stable rank $\mathrm{R}(\cdot)$ is a more robust alternative to the standard rank in that it is less sensitive to small changes in the spectrum. Supposing that $P \geq M$, two extreme cases are worth noting here. First, if the rows of A are orthonormal, then $\mathrm{R}(A) = M$. In this case, the rows of A are equal in length and have "diverse" directions. Second, if the rows of A are identical, then $\mathrm{R}(A) = 1$.

The star of this show will in fact be the stable rank of the attractor $\mathbb{A}$, which we define to be

$$\mathrm{R}_{H,T,M}(\mathbb{A}) := \inf_{x,y \in \mathbb{A},\ x \neq y} \mathrm{R}\left(X_{H,T,M} - Y_{H,T,M}\right), \tag{16}$$

where $X_{H,T,M} \in \mathbb{R}^{M \times P}$ is defined in (11) (see also (12)), and $Y_{H,T,M}$ is defined analogously with y in place of x. If $P \geq M$, then (15) dictates that

$$\mathrm{R}_{H,T,M}(\mathbb{A}) \in [1, M]. \tag{17}$$

For example, when $H(\mathbb{A})$ is a subset of an r-dimensional subspace (say with $r \ll M$), then (16) dictates that $\mathrm{R}_{H,T,M}(\mathbb{A}) \leq r \ll M$. On the other hand, if the rows of $X_{H,T,M} - Y_{H,T,M}$ have similar lengths and diverse directions (for every $x, y \in \mathbb{A}$), then $\mathrm{R}_{H,T,M}(\mathbb{A})$ might be close to M. As we see next, the larger $\mathrm{R}_{H,T,M}(\mathbb{A})$, the better.

C. Main Result

We are now in position to state the main result of this work.

Theorem III.1. (Stable Takens' embedding theorem) *Let $\mathbb{A} \subset \mathbb{R}^N$ be a smooth, bounded, and boundary-less submanifold of $\mathbb{R}^N$. For a fixed sampling interval $T > 0$, assume that $\phi_T(\cdot)$ is a diffeomorphism on $\mathbb{A}$ and that the singular values of the derivative of $\phi_T(\cdot)$ belong to the interval $[\sigma_{\min}, \sigma_{\max}] \subset (0, \infty)$. For an integer P, fix the measurement basis functions $h_p : \mathbb{A} \to \mathbb{R}$ for $p \in \{1, 2, \ldots, P\}$ and let $\mathcal{H}$ be the linear span of $\{h_p(\cdot)\}_p$. The random coefficient vector $\alpha \in \mathbb{R}^P$ is assumed to have entries that are i.i.d. zero-mean and unit-variance subgaussian random variables with subgaussian norm θ. A random variable Z is subgaussian if its subgaussian norm $\|Z\|_{\psi_2}$ is finite, where $\|Z\|_{\psi_2} := \sup_{p \geq 1} (\mathbb{E}|Z|^p)^{\frac{1}{p}}$. Qualitatively speaking, the tail of (the distribution of) a subgaussian random variable is similar to that of a Gaussian random variable, hence the name [64]. In particular, Gaussian random variables are subgaussian.*

15

Consider the map $H : \mathbb{A} \to \mathbb{R}^P$ constructed in (8), and suppose that $H(\cdot)$ satisfies the assumptions A1–A3 listed in Section III A. Let $\mathrm{R}_{H,T,M}(\mathbb{A})$ denote the stable-rank of $\mathbb{A}$ as defined in (16). For arbitrary isometry constant $\delta \in (0,1)$ and failure probability $\rho \in (0,1)$, suppose that

$$
\mathrm{R}_{H,T,M}(\mathbb{A}) \geq C_\theta \max \left[\delta^{-2} \dim(\mathbb{A}) \log \left(\eta_{\max} \sqrt{\dim(\mathbb{A})} \left(\frac{\sigma_{\min}^{-2M\dim(\mathbb{A})} - 1}{\sigma_{\min}^{-2\dim(\mathbb{A})} - 1} \right)^{\frac{1}{2\dim(\mathbb{A})}} \frac{\mathrm{vol}\,(\mathbb{A})^{\frac{1}{\dim(\mathbb{A})}}}{\mathrm{rch}(\mathbb{A}_{H,T,M})} \right) \right.
$$
$$
\left. , e^{-\min W\left(\frac{-\delta^2}{\dim(\mathbb{A})} \right)}, \delta^{-2} \log \left(\frac{1}{\rho} \right) \right] \tag{18}
$$

where C_θ is a constant that depends only on θ, and make the mild assumption that

$$
\frac{\mathrm{vol}\,(\mathbb{A})^{\frac{1}{\dim(\mathbb{A})}}}{\mathrm{rch}\,(\mathbb{A}_{H,T,M})} \gtrsim \frac{1}{\eta_{\min} \sqrt{\dim(\mathbb{A})}} \left(\frac{\sigma_{\max}^{-2M\dim(\mathbb{A})} - 1}{\sigma_{\max}^{-2\dim(\mathbb{A})} - 1} \right)^{-\frac{1}{2\dim(\mathbb{A})}}, \tag{19}
$$

with $\mathbb{A}_{H,T,M} \subset \mathbb{R}^{MP}$ defined in (13) and $W(\cdot)$ denoting the Lambert W-function. See Appendix A or [65, §4.13] for the exact definition of the Lambert W-function. Very roughly, the exponential term involving the Lambert W-function in (18) scales like $\dim(\mathbb{A}) \log(\dim(\mathbb{A}))$. This assumption requires the volume of $\mathbb{A}$ not to be too small. Similar assumptions have appeared in earlier works (e.g., [33]).

Then, except with a probability of at most ρ (over the choice of α), the delay-coordinate map $F_{h_\alpha,T,M}(\cdot)$ stably embeds $\mathbb{A}$ in the sense that

$$
(1-\delta)l_H^2 \,\mathrm{geo}(\mathbb{A})^{-2} \frac{\sigma_{\max}^{-2M} - 1}{M(\sigma_{\max}^{-2} - 1)} \leq \frac{\|F_{h_\alpha,T,M}(x) - F_{h_\alpha,T,M}(y)\|_2^2}{M\|x-y\|_2^2} \leq (1+\delta)u_H^2 \,\mathrm{geo}(\mathbb{A})^2 \frac{\sigma_{\min}^{-2M} - 1}{M(\sigma_{\min}^{-2} - 1)}, \tag{20}
$$

for every pair $x, y \in \mathbb{A}$ with $x \neq y$.

The proof of this result is found in Appendix A. In (20), regarding the behavior of the terms involving $\sigma_{\max}$ and $\sigma_{\min}$ (the largest and smallest singular values of the derivative of $\phi_T(\cdot)$), we note that these terms are close to 1 if the singular values cluster near 1. In particular,

$$
\lim_{\sigma \to 1} \frac{\sigma^{-2M} - 1}{M(\sigma^{-2} - 1)} = 1.
$$

D. Observations and Interpretation

Several remarks are in order to help shape our understanding of Theorem III.1.

Remark III.1. (**Comparison with Takens' theorem**) Let us fix the measurement basis functions $\{h_p(\cdot)\}$. Note that the distribution of the random coefficient vector $\alpha \in \mathbb{R}^P$ in Theorem III.1 induces a distribution on the space of measurement functions, $\mathcal{H} = \mathrm{span}[\{h_p(\cdot)\}]$. Qualitatively speaking, Theorem III.1 establishes that, except on a subset with an exponentially small measure, every function in $\mathcal{H}$ forms a delay-coordinate map that stably embeds $\mathbb{A}$, if $\mathrm{R}_{H,T,M}(\mathbb{A})$ is proportional to $\dim(\mathbb{A})$ with a proportionality constant that depends chiefly on the geometry of $\mathbb{A}$ and the flow $\phi_T(\cdot)$.

In contrast, Takens' original theorem (Theorem II.1) established that generic choices of the flow ϕ_T and measurement function h will yield an embedding so long as that $M > 2\dim(\mathbb{A})$. The refinement of Takens' theorem by Sauer et al. [15] established that, for a fixed flow ϕ_T and a random choice of h from a certain probe space, one will obtain an embedding with probability one. This result also required that $M > 2\dim(\mathbb{A})$ but placed certain restrictions on the periodicities of the orbits of ϕ_T on $\mathbb{A}$.

Thus, Theorem III.1 provides a stronger embedding guarantee than the topological and probabilistic Takens' theorems, but it does so with a nonzero failure probability and it is contingent on a condition involving the stable rank $\mathrm{R}_{H,T,M}(\mathbb{A})$. If this condition can be satisfied for a given attractor $\mathbb{A}$, flow ϕ_T, and space of measurement functions $\mathcal{H}$, it may require choosing the number of delays M larger than $2\dim(\mathbb{A})$.

Remark III.2. (**Stable rank**) The requirement on the stable rank of $\mathbb{A}$ in (18) merits special attention. Let us fix the measurement basis functions (and consequently the map $H(\cdot)$). The condition in (18) must be satisfied to stably embed $\mathbb{A}$, which may require the user to sufficiently increase $\mathrm{R}_{H,T,M}(\mathbb{A})$ by adjusting the sampling interval T and the number of delays M. In fact, (18) helps justify certain design rules that are commonly employed in constructing delay-coordinate mappings.

Suppose for the moment that an oracle could inform the user of $\mathrm{R}_{H,T,M}(\mathbb{A})$ for a given pair (T, M) and let us examine the behavior of the stable rank under these variables. If $P < M$, then $\mathrm{R}_{H,T,M}(\mathbb{A})$ is upper bounded by P. However, if $P \geq M$, recall from (17) that $\mathrm{R}_{H,T,M}(\mathbb{A}) \in [1, M]$. If $\mathrm{R}_{H,T,M}(\mathbb{A}) \approx M$, the user could eventually enforce (18) by increasing M (thereby stably embedding $\mathbb{A}$). But how can the user enforce $\mathrm{R}_{H,T,M}(\mathbb{A}) \approx M$ by adjusting T? From (16), $\mathrm{R}_{H,T,M}(\mathbb{A}) \approx M$ means that the rows of $X_{H,T,M} - Y_{H,T,M} \in \mathbb{R}^{M \times P}$ are nearly orthonormal for every pair $x, y \in \mathbb{A}$ (see the discussion in Section III B). Roughly speaking, the following considerations are relevant:

- For the rows of $X_{H,T,M} - Y_{H,T,M}$ (see (12)) to have nearly the same length, T must be substantially smaller (in magnitude) than the Lyapunov exponents of the flow $\phi_T(\cdot)$ on $\mathbb{A}$ so that

$$\|\phi_{H,T}^{-m}(H(x)) - \phi_{H,T}^{-m}(H(y))\|_2 \approx \|H(x) - H(y)\|_2, \qquad \forall x, y \in \mathbb{A}, \, m \in \{0, 1, \ldots, M-1\}, \tag{21}$$

by the invariance of Lyapunov exponents under the diffeomorphism $H(\cdot)$ [11, Section 11.2]. Note that when T is comparable to the Lyapunov exponents (in magnitude), then the rows of $X_{H,T,M} - Y_{H,T,M}$ might differ considerably in length, and $\mathrm{R}_{H,T,M}(\mathbb{A})$ is likely to be small (leading to a poor embedding of $\mathbb{A}$).

  This aspect of our theoretical result mirrors the well-recognized phenomenon of *irrelevancy* [16, 19, 20]. Indeed, when T is comparable to the Lyapunov exponents (in magnitude), entries of the delay vector $F_{h_\alpha,T,M}(x(t)) \in \mathbb{R}^M$ are likely to be "causally independent." In this case, the trajectory $F_{h_\alpha,T,M}(x(\cdot)) \subset \mathbb{R}^M$ in the reconstruction space will be unnecessarily more complex than the original trajectory $x(\cdot) \subset \mathbb{R}^N$ in the state space.

- For the rows of $X_{H,T,M} - Y_{H,T,M}$ to be nearly orthogonal for every $x, y \in \mathbb{A}$, the trajectories of the flow $\phi_{H,T}(\cdot)$ on $H(\mathbb{A})$ should be "diverse" in that they should "visit" different dimensions as time progresses. Adjusting T here might help push the rows of $X_{H,T,M} - Y_{H,T,M}$ to become nearly orthogonal. However, when T is very small, the rows of $X_{H,T,M} - Y_{H,T,M}$ (for a pair $x, y \in \mathbb{A}$) are similar in direction and length, and consequently $\mathrm{R}_{H,T,M}(\mathbb{A})$ is likely to be small (resulting in a poor embedding).

  Similarly, this aspect of our theoretical result mirrors a known phenomenon called *redundancy* [16, 19, 20]. Indeed, when T is very small, the adjacent entries of a delay vector $F_{h_\alpha,T,M}(x(t)) \in \mathbb{R}^M$ are likely to be highly similar (or "correlated") and the information contained in $F_{h_\alpha,T,M}(x(t))$ is largely redundant. In this case, the trajectory $F_{h_\alpha,T,M}(x(\cdot)) \subset \mathbb{R}^M$ in the reconstruction space will be stretched out along the identity line (regardless of the geometry of the attractor $\mathbb{A} \subset \mathbb{R}^N$).

To summarize the main points, if T is chosen too large or too small, then $\mathrm{R}_{H,T,M}(\mathbb{A})$ will rapidly plateau when the user increases M. Thus, our theoretical findings echo the (mainly heuristically-investigated) trade-off between irrelevancy and redundancy in the literature,

suggesting the user may improve the embedding quality if a sampling interval in this ideal intermediate range can be found.

Remark III.3. (**Choice of T and M**) The discussion in Remark III.2 raises the following question:

- Can the user experimentally find the right range for T and M without prior knowledge of the quantities involved in (18)?

To answer this question, we first point out that a similar issue has arisen in the past with choosing the number of delays M for Takens' original theorem. As a practical method for setting this parameter [11], the community has observed that $\dim(\mathbb{A})$ is preserved under delay-coordinate mapping effectively *as long as $M > 2\dim(\mathbb{A})$* and there is no noise. This observation suggests the following procedure for estimating $\dim(\mathbb{A})$ and consequently estimating the required number of delays M in Takens' theorem. For fixed T and every M within a fixed range $\{M_1, \ldots, M_2\}$, the user constructs a sequence of delay-coordinate maps for many example test observations from the system. For each M, the user applies the Grassberger-Procaccia algorithm [66] to estimate $\dim(F_{h,T,M}(\mathbb{A}))$ and searches for a range of values of M where the graph of $\dim(F_{h,T,M}(\mathbb{A}))$ (versus M) plateaus. This plateau is an estimate for $\dim(\mathbb{A})$, and a reasonable choice of M immediately follows. When noise is present, this plateau may disappear at large values of M as well, resulting in a "sweet spot" in the graph where $M = O(2\dim(\mathbb{A}))$ and $\dim(F_{h,T,M}(\mathbb{A})) = \dim(\mathbb{A})$.

Returning to the present problem, a similar approach can be used. If (18) indeed holds, then (with high probability) the delay-coordinate map with parameters T and M stably embeds $\mathbb{A}$ into the reconstruction space $\mathbb{R}^M$ and the volume of $\mathbb{A}$ is preserved. In general, (2) implies that

$$(\epsilon_l M)^{\frac{\dim(\mathbb{A})}{2}} \mathrm{vol}(\mathbb{A}) \leq \mathrm{vol}\left(F_{h_\alpha,T,M}(\mathbb{A})\right) \leq (\epsilon_u M)^{\frac{\dim(\mathbb{A})}{2}} \mathrm{vol}(\mathbb{A}). \tag{22}$$

This claim is proved similar to those in Appendix B. This observation implies a variant of the algorithm described above where volume is used in place of dimension to find the correct range of T and M, which we detail in Table I.

Remark III.4. (**Quality of embedding**) Let us again fix the basis functions (and thus $H(\cdot)$), and suppose that (18) holds for a given isometry constant $\delta \in (0,1)$ and failure probability $\rho \in (0,1)$. Then the quality of embedding in (20) clearly depends on

TABLE I. *Prescription to find the proper range of the sampling interval T and the number of delays M in delay-coordinate mapping.*

How to choose T and M in delay-coordinate mapping

1. Given a time-series $\{s_i\}_i$ and a scalar measurement function $h(\cdot)$, compute the delay vectors $\{F_{h,T,M}(x(iT))\}_i \subset \mathbb{R}^M$ for every pair (T, M) in the window $[T_{\min}, T_{\max}] \times [M_{\min} : M_{\max}]$.

2. For each pair (T, M), empirically compute the dimension $d_{T,M}$ and volume $V_{T,M}$ of the surface formed by the delay vectors $\{F_{h,T,M}(x(iT))\}_i$, and plot $\frac{V_{T,M}}{\sqrt{M^{d_{T,M}}}}$ for various pairs (T, M) in the above window.

3. Find the range of (T, M) for which the graph is nearly constant. This provides the recommended range for T and M in delay-coordinate mapping of the system under study.

- the bi-Lipschitz constants of $H(\cdot)$ (i.e., l_H, u_H);

- the spectrum of the derivative of the flow $\phi_T(\cdot)$ (through $\sigma_{\min}, \sigma_{\max}$); and

- the geodesic regularity of the attractor $\mathbb{A}$ (i.e., $\text{geo}(\mathbb{A})$).

Large values of $\frac{u_H}{l_H}$, $\frac{\sigma_{\max}}{\sigma_{\min}}$, and $\text{geo}(\mathbb{A})$ in (20) all result in a poor embedding guarantee for $\mathbb{A}$ (i.e., a large disparity between the upper and lower bounds in (20)). In particular, when the dynamical system is highly unpredictable (e.g., has a large Lyapunov exponent), then $\frac{\sigma_{\max}}{\sigma_{\min}}$ is likely to be very large and the embedding guarantee (and, indeed, the embedding itself) is likely to be poor. In a nutshell, stably embedding unpredictable systems (e.g., chaotic systems) is often difficult.

Remark III.5. **(Orbits and other pathologies)** The flow $\phi_T(\cdot)$ has an orbit with period n if $\phi_T^n(x) = \phi_{nT}(x) = x$ for some $x \in \mathbb{A}$. As noted in Remark III.1, the probabilistic statement of Takens' theorem by Sauer et al. [15] placed certain restrictions on the periodicities of the orbits of ϕ_T. Indeed, the existence of orbits also typically deteriorates the stable rank of a system. As an extreme example, consider an orbit of period one, otherwise known as a *fixed*

point: $\phi_T(x) = x$ for some $x \in \mathbb{A}$. Using (16), we may easily verify that $\mathrm{R}_{H,T,M}(\mathbb{A}) = 1$ for any choice of basis functions and any number of delays M. That is, the stable rank of $\mathbb{A}$ does not increase at all when the user increases M. In view of (18), this leads to a very poor embedding of the attractor $\mathbb{A}$. We note that orbits of period one are explicitly forbidden by Sauer et al. [15] and implicitly forbidden in Theorem II.1 through the genericity of ϕ_T.

E. Extensions to Strange Attractors

While our discussion thus far has focused on attractors that comprise smooth submanifolds of $\mathbb{R}^n$, many dissipative dynamical systems (e.g., chaotic systems) converge onto attractors that are not smooth submanifolds of the Euclidean space. In this section, we discuss what changes when considering the stable embedding of more general (e.g., strange) attractors. In what follows, we continue to assume that the state lies on the attractor $\mathbb{A}$ so that for every time t, $x(t) \in \mathbb{A}$.

1. Global enveloping manifolds

The easiest scenario arises when there exists a *global enveloping manifold* $\mathbb{M}$ that subsumes the attractor $\mathbb{A}$. Roughly speaking, we say that $\mathbb{M} \subset \mathbb{R}^n$ is a global enveloping manifold of an attractor $\mathbb{A} \subset \mathbb{R}^n$ if $\mathbb{A} \subset \mathbb{M}$ and at every point $x \in \mathbb{A}$, $\mathbb{T}_x \mathbb{A} = \mathbb{T}_x \mathbb{M}$ (see [67, 68] for a more precise definition). Here, $\mathbb{T}_x \mathbb{M}$ denotes the conventional tangent space of $\mathbb{M}$ at x (recall Section II C), and $\mathbb{T}_x \mathbb{A}$ denotes a generalized tangent space of $\mathbb{A}$ at x, defined as follows.

Definition III.1. (**Generalized tangent space [68]**) *Consider an attractor $\mathbb{A} \subset \mathbb{R}^N$ and a point $x \in \mathbb{A}$. The generalized tangent space of $\mathbb{A}$ at x, denoted $\mathbb{T}_x \mathbb{A}$, is the smallest linear space containing all unit vectors of the form $(z_i - y_i)/\|z_i - y_i\|_2$ generated by sequences $\{y_i\}$ and $\{z_i\}$ in $\mathbb{A}$ with $y_i \to x$ and $z_i \to x$.*

In scenarios where there does exist a global enveloping manifold $\mathbb{M}$ for $\mathbb{A}$, Theorem III.1 can be naturally extended to provide conditions for the stable embedding of $\mathbb{M}$ (and thus $\mathbb{A}$). In order to prove this result, one merely replaces $\mathbb{A}$ with $\mathbb{M}$ throughout the statement and the proof of Theorem III.1; consequently, all of the geometric quantities that appear in the resulting bound—dimension, volume, reach, and so on—will refer to $\mathbb{M}$ instead of $\mathbb{A}$. As

noted in Remark III.4, when the dynamical system is highly unpredictable (e.g., has a large Lyapunov exponent), then $\frac{\sigma_{\max}}{\sigma_{\min}}$ is likely to be very large and the embedding guarantee (and, indeed, the embedding itself) is likely to be poor. In such a case, Remark A.1 may have some value. However, because it may be unreasonable to assume that the enveloping manifold is invariant under the flow (i.e., that $\phi_T(\mathbb{M}) = \mathbb{M}$), one may relax this assumption in the statement and proof of the theorem; all that is needed is that ϕ_T acts as a diffeomorphism between $\mathbb{M}$ and $\phi_T(\mathbb{M})$ (or, more precisely, between $\mathbb{M}$ and each of $\phi_T^{-1}(\mathbb{M}), \ldots, \phi_T^{-M+1}(\mathbb{M})$), and that the assumptions on H listed in Section III A hold not only on $\mathbb{M}$ but also on each of $\phi_T^{-1}(\mathbb{M}), \ldots, \phi_T^{-M+1}(\mathbb{M})$.

2.   *More general attractors*

Alas, a counter-example in [68] shows that not all subsets of Euclidean space, and thus potentially not all attractors of dynamical systems, can have a global enveloping manifold. When the attractor $\mathbb{A}$ does not have a global enveloping manifold, we require a few additional definitions that will endow the attractor with certain geometric properties that make it amenable for our analysis. We first describe these properties in terms of a general subset $\mathbb{B} \subset \mathbb{R}^N$.

Definition III.2. (**Box-counting dimension [15]**) *Consider a set $\mathbb{B} \subset \mathbb{R}^N$. Suppose $\mathbb{R}^N$ is divided into cubes of size ζ by a grid based at points whose coordinates are ζ-multiples of the integers. Let $\mathcal{N}(\zeta)$ be the number of boxes or cubes of size ζ that intersect $\mathbb{B}$. Then the box-counting dimension of $\mathbb{B}$, denoted by $\mathrm{boxdim}(\mathbb{B})$, is defined as*

$$\mathrm{boxdim}(\mathbb{B}) := \lim_{\zeta \to 0} -\frac{\log \mathcal{N}(\zeta)}{\log(\zeta)}.$$

Definition III.3. (**Covering regularity**) *We say that the set $\mathbb{B}$ has covering regularity $\mathrm{cov}(\mathbb{B})$ (depending on some maximum size ζ_0) if for every $\zeta \leq \zeta_0$,*

$$\mathcal{N}(\zeta) \leq \mathrm{cov}(\mathbb{B})\zeta^{-\mathrm{boxdim}(\mathbb{B})},$$

where $\mathcal{N}(\zeta)$ is the number of boxes or cubes of size ζ that intersect $\mathbb{B}$ (see Definition III.2).

One can think of the covering regularity $\mathrm{cov}(\mathbb{B})$ as a proxy for the volume of $\mathbb{B}$ because volume is proportional to $\mathcal{N}(\zeta)\zeta^{\mathrm{boxdim}(\mathbb{B})}$ in the limit of small ζ when $\mathbb{B}$ is a submanifold.

Definition III.4. **(Tangent covering regularity)** *We say that the set* $\mathbb{B}$ *has tangent cover-ing regularity* $\mathrm{tancov}(\mathbb{B})$ *(depending on some maximum size* ζ_0*) if for every* $a \in \mathbb{B}$*, whenever* $\|x - a\|_2, \|y - a\|_2 \leq \zeta \leq \zeta_0$ *for some* $x, y \in \mathbb{B}$*, we can find a* $v \in \mathbb{T}_a\mathbb{B}$ *such that*

$$\left\| v - \frac{x - y}{\|x - y\|_2} \right\|_2 \leq \mathrm{tancov}(\mathbb{B})\zeta.$$

Here $\mathrm{tancov}(\mathbb{B})$ can be thought of as a measure the curvature of $\mathbb{B}$.

Definition III.5. **(Tangent dimension)** *We define the tangent dimension* $\mathrm{tandim}(\mathbb{B})$ *of the set* $\mathbb{B}$ *as*

$$\mathrm{tandim}(\mathbb{B}) := \sup_{x \in \mathbb{B}} \dim\left(\mathbb{T}_x\mathbb{B}\right),$$

where $\mathbb{T}_x\mathbb{B}$ *refers to the generalized tangent space of* $\mathbb{B}$ *at* x *(see Definition III.1).*

As noted in [68], the tangent dimension bounds the box-counting dimension from above: for any set $\mathbb{B}$, $\mathrm{tandim}(\mathbb{B}) \geq \mathrm{boxdim}(\mathbb{B})$. In what follows, we shall ignore the dependence of the regularity quantities on their maximal resolution ζ_0.

With these properties thus defined, we present our result for the stable embedding of a general (including strange) attractor. The following theorem makes a series of assumptions not on the attractor $\mathbb{A}$ itself, but rather on the trajectory attractor $\mathbb{A}_{H,T,M} \subset \mathbb{R}^{MP}$ defined in (13). We discuss these conditions further after presenting the main result, which is proved in Appendix C.

Theorem III.2. **(Stable Takens' embedding theorem for strange attractors)** *Let* $\mathbb{A} \subset \mathbb{R}^N$ *be an attractor. For a fixed sampling interval* $T > 0$*, assume that* $\phi_T(\cdot)$ *is a flow on* $\mathbb{A}$*. For an integer* P*, fix the measurement basis functions* $h_p : \mathbb{A} \to \mathbb{R}$ *for* $p \in \{1, 2, \ldots, P\}$ *and let* $\mathcal{H}$ *be the linear span of* $\{h_p(\cdot)\}_p$*. The random coefficient vector* $\alpha \in \mathbb{R}^P$ *is assumed to have entries that are i.i.d. zero-mean and unit-variance subgaussian random variables with subgaussian norm* θ*.*

Consider the map $H : \mathbb{A} \to \mathbb{R}^P$ *constructed in (8), and suppose that* $H(\cdot)$ *satisfies as-sumption A1 listed in Section III A. Let* $\mathbb{A}_{H,T,M} \subset \mathbb{R}^{MP}$ *be the associated trajectory attrac-tor defined in (13). Suppose* $\mathbb{A}_{H,T,M}$ *has box-counting dimension* $\mathrm{boxdim}(\mathbb{A}_{H,T,M})$*, tangent dimension* $\mathrm{tandim}(\mathbb{A}_{H,T,M})$*, covering regularity* $\mathrm{cov}(\mathbb{A}_{H,T,M}) > 1$*, and tangent covering regu-larity* $\mathrm{tancov}(\mathbb{A}_{H,T,M}) > \frac{3}{\sqrt{MP}}$*. Finally, let* $\mathrm{R}_{H,T,M}(\mathbb{A})$ *denote the stable-rank of* $\mathbb{A}$ *as defined in (16).*

For arbitrary isometry constant $\delta \in (0,1)$ and failure probability $\rho \in (0,1)$, suppose that

$$\mathrm{R}_{H,T,M}(\mathbb{A}) \geq$$

$$C'_\theta \max \left[\delta^{-2} \operatorname{tandim}(\mathbb{A}_{H,T,M}) \log \left(\sqrt{MP} \operatorname{tancov}(\mathbb{A}_{H,T,M}) \left(\operatorname{cov}(\mathbb{A}_{H,T,M}) \right)^{1/\operatorname{boxdim}(\mathbb{A}_{H,T,M})} \right), \right.$$

$$\left. e^{-\min W\left(\frac{-\delta^2}{\operatorname{tandim}(\mathbb{A}_{H,T,M})} \right)}, \delta^{-2} \log \left(\frac{1}{\rho} \right) \right] \tag{23}$$

where C'_θ is a constant that depends only on θ. Then except with a probability of at most ρ (over the choice of α),

$$1 - \delta \leq \frac{\|F_{h_\alpha,T,M}(x) - F_{h_\alpha,T,M}(y)\|_F^2}{\sum_{m=0}^{M-1} \|H\left(\phi_T^{-m}(x)\right) - H\left(\phi_T^{-m}(y)\right)\|_2^2} \leq 1 + \delta \tag{24}$$

holds for all $x, y \in \mathbb{A}$ with $x \neq y$.

Moreover, if (24) holds for all $x, y \in \mathbb{A}$ with $x \neq y$ and if there exist quantities $\operatorname{geo}(\mathbb{A})$, $\sigma_{\min}$, $\sigma_{\max}$ such that for all $x, y \in \mathbb{A}$ and $m = 1, 2, \ldots, M-1$,

$$\operatorname{geo}(\mathbb{A})^{-1} \sigma_{\max}^{-m} \|x - y\|_2 \leq \|\phi_T^{-m}(x) - \phi_T^{-m}(y)\|_2 \leq \operatorname{geo}(\mathbb{A}) \sigma_{\min}^{-m} \|x - y\|_2, \tag{25}$$

it follows that

$$(1-\delta) l_H^2 \operatorname{geo}(\mathbb{A})^{-2} \frac{\sigma_{\max}^{-2M} - 1}{\sigma_{\max}^{-2} - 1} \leq \frac{\|F_{h_\alpha,T,M}(x) - F_{h_\alpha,T,M}(y)\|_2^2}{\|x-y\|_2^2} \leq (1+\delta) u_H^2 \operatorname{geo}(\mathbb{A})^2 \frac{\sigma_{\min}^{-2M} - 1}{\sigma_{\min}^{-2} - 1} \tag{26}$$

holds for all $x, y \in \mathbb{A}$ with $x \neq y$.

Much like our original Theorem III.1, Theorem III.2 guarantees a stable embedding of an attractor with high probability, under the condition that the stable rank $\mathrm{R}_{H,T,M}(\mathbb{A})$ is sufficiently large. However, whereas the right hand side of (18) involves mostly properties of the attractor $\mathbb{A}$ itself, the right hand side of (23) references properties of the trajectory attractor $\mathbb{A}_{H,T,M}$ instead. Indeed, a key step in the proof is bounding the covering number of the set of all normalized secants of the trajectory attractor. In the proof of Theorem III.1, we used Lemma A.1 to relate properties of $\mathbb{A}_{H,T,M}$ to those of $\mathbb{A}$. In the case of strange attractors, we leave this connection as an open question.

Perhaps interestingly, Theorem III.2 does not require any assumptions regarding ϕ_T or H being a diffeomorphism, nor any assumptions on the singular values of their derivatives. Such properties do likely affect the quality of the embedding. In the proof of Theorem III.1,

we used these properties both in Lemma A.1 (to relate properties of $\mathbb{A}_{H,T,M}$ to those of $\mathbb{A}$) and to guarantee that a condition equivalent to (25) holds. However, the original proof of the condition equivalent to (25) required an argument involving geodesic distance, which is not appropriate for a strange attractor. Here, we pull out (25) as its own assumption, which could conceivably hold even for a strange attractor. Thus, in (25), the quantities geo($\mathbb{A}$), $\sigma_{\min}$, $\sigma_{\max}$ do not necessarily refer to the geodesic regularity of the attractor $\mathbb{A}$ or the singular values of the derivative of $\phi_T(\cdot)$. However, for (25) to hold, these parameters would likely play similar roles to those that they played in Theorem III.1.

IV.   EXAMPLES

In this section, we present two examples that support the theoretical findings in Section III C, emphasizing the relationship between the stable rank of a system and the number of delays in delay-coordinate mapping.

A.   Moment Curve

We begin with an example where we can analytically calculate (or bound) the quantities of interest. Strictly speaking, Theorem III.1 applies to subsets of $\mathbb{R}^N$ and not to $\mathbb{A} \subset \mathbb{C}^N$ as in this example. However, study of the "real" moment curve (formed from the real part of $\gamma(\cdot)$) is far more tedious, and is therefore not pursued here for the sake of the clarity. In fact, we strongly suspect that Lemma 15 in [33] and consequently Theorem III.1 can be extended (with minor changes) to account for complex attractors.

For an integer N, let $\mathbb{A}$ be the moment curve in $\mathbb{C}^N$. That is,

$$\mathbb{A} = \{\gamma(t) : t \geq 0\} \subset \mathbb{C}^N, \quad \gamma(t) = \begin{bmatrix} 1 \\ e^{\mathrm{i}2\pi t} \\ \vdots \\ e^{\mathrm{i}2\pi(N-1)t} \end{bmatrix}. \tag{27}$$

Note that $\mathbb{A}$ is a closed curve because $\gamma(n) = \gamma(0)$, for every integer n. For a fixed $T > 0$, we endow $\mathbb{A}$ with a linear dynamical system with flow $\phi_T(\cdot)$. This linear flow, which we identify with an $N \times N$ matrix, is specified as

$$\phi_T = \mathrm{diag}\,[\gamma(T)] \in \mathbb{C}^{N \times N}, \tag{28}$$

where diag[a] returns the diagonal matrix formed from the entries of vector a. For any $t \geq 0$, observe that $\phi_T(\gamma(t)) = \phi_T \gamma(t) = \gamma(t + T)$; that is $\mathbb{A} = \gamma(\cdot)$ is parametrized by time.

Let $\mathcal{H}$ be the space of all linear functionals on $\mathbb{C}^N$, so that every scalar measurement function may be characterized as $h_\alpha(\cdot) = \langle \cdot, \alpha \rangle$ for some $\alpha \in \mathbb{C}^N$. In the language of Theorem III.1, we set $P = N$ and take $H(\cdot)$ to be the identity operator (and, in particular, $H(\mathbb{A}) = \mathbb{A}$). Assume also that the entries of α are independent Gaussian random variables with zero mean and unit variance (the variance of a complex random variable is the sum of the variances of its real and imaginary parts).

We next compute the relevant geometric quantities. Since $\mathbb{A}$ is a curve, $\dim(\mathbb{A}) = 1$, and $\mathrm{vol}(\mathbb{A})$ is simply its length:

$$
\begin{aligned}
\mathrm{vol}(\mathbb{A}) = \mathrm{length}(\gamma(\cdot)) &= \int_0^1 \left\| \frac{d\gamma(t)}{dt} \right\|_2 dt \\
&= 2\pi \sqrt{\sum_{n=0}^{N-1} n^2} \int_0^1 dt \qquad \text{(see (27))} \\
&= \pi \sqrt{\frac{2}{3}(N-1)N(2N-1)}.
\end{aligned}
\tag{29}
$$

Next, we turn to the geodesic regularity of the moment curve which involves comparing geodesic and Euclidean distances between an arbitrary pair of points on $\mathbb{A}$. Using (29) (and the implicit observation therein that $\gamma(\cdot)$ has constant "speed"), we deduce that the geodesic distance between $\gamma(t_1), \gamma(t_2) \in \mathbb{A}$ is given by

$$
\begin{aligned}
d_\mathbb{A}(\gamma(t_1), \gamma(t_2)) &= |t_1 - t_2| \, \mathrm{length}(\gamma(\cdot)) \\
&= |t_1 - t_2| \, \pi \sqrt{\frac{2}{3}(N-1)N(2N-1)}, \qquad \forall t_1, t_2 \in [0, 1).
\end{aligned}
\tag{30}
$$

In Appendix E, we calculate the Euclidean distance $\|\gamma(t_1) - \gamma(t_2)\|_2$ and estimate the geodesic regularity of the moment curve by comparing the two metrics.

Lemma IV.1. **(Geodesic regularity)** *For an integer N, let $\mathbb{A}$ be the moment curve in $\mathbb{C}^N$ (see (27)). Then, the geodesic regularity of $\mathbb{A}$ (see 5) is bounded as*

$$
\mathrm{geo}(\mathbb{A}) \leq \frac{2\pi^2}{3(1 - \beta_1)} N(N-1), \qquad \forall N > N_m.
\tag{31}
$$

Above, $\beta_1 > 0$ is a (small) absolute constant, and N_m is a sufficiently large integer (which is subject to change in every appearance).

In other words, the geodesic regularity of $\mathbb{A}$ is poor, $\text{geo}(\mathbb{A}) = O(N^2)$ which, in light of (20), suggests that delay-coordinate mapping might poorly embed this system (when the dimension of the state space N is large). The guarantees in Theorem III.1 appear to be conservative here as our simulations indicate later in this section.

Next, to compute the reach, we borrow from Lemma 1 in [33]:

$$\beta_2 \sqrt{N} \le \text{rch}(\mathbb{A}) \le \sqrt{N}, \qquad \forall N > N_m. \tag{32}$$

Here, $\beta_2 < 1$ is independent of N, and N_m is a sufficiently large integer (note that the discrepancy in the definition of moment curve here and in [33] is inconsequential). That is, fortuitously, the reach of the moment curve is relatively large. Next, we turn to the stable rank of this system. The estimate below is obtained in Appendix F.

Lemma IV.2. *For an integer N, let $\mathbb{A}$ be the moment curve in $\mathbb{C}^N$ (see (27)). For $T \in (0, \frac{1}{M}]$, equip $\mathbb{A}$ with the linear flow $\phi_T(\cdot)$ specified in (28). Then, the stable rank of $\mathbb{A}$ (as defined in (16)) satisfies*

$$\frac{M}{20 + \frac{40}{N \sin(\pi T)} \log \left(e / \tan \left(\frac{\pi T}{2}\right)\right)} \le \text{R}_{H,T,M}(\mathbb{A}) \le M, \qquad \forall N > N_m, \tag{33}$$

where N_m is a sufficiently large integer.

Roughly speaking, as long as $\frac{1}{N} \lesssim T \le \frac{1}{M}$, the stable rank of our system is large ($\text{R}_{H,T,M}(\mathbb{A}) \approx M$). (Note that $M \lesssim N$ is necessary for this claim to hold.) Since the stable rank of any system is bounded by the number of delays M (see (17)), this result is nearly ideal.

Let us now empirically compute the stable rank of this system (see (16)) for variable number of delays M and with $N = 250$ and $T = 1/(N + \xi)$, where ξ is chosen randomly from a standard normal distribution. The outcome appears in Figure 2(a). To see the connection between the stable rank and the quality of embedding, we plot in Figure 2(b) the isometry constants $\epsilon_l \le \epsilon_u$ (recall (2)) versus the number of delays M. To produce the plot, we generated 100 independent copies of $h_\alpha(\cdot)$ and computed the isometry constants according to (2). The curve shows the mean isometry constants (over 100 repetitions). As M increases, the stable rank increases (improves) and the isometry constants tighten (the quality of embedding improves); this matches Theorem III.1.
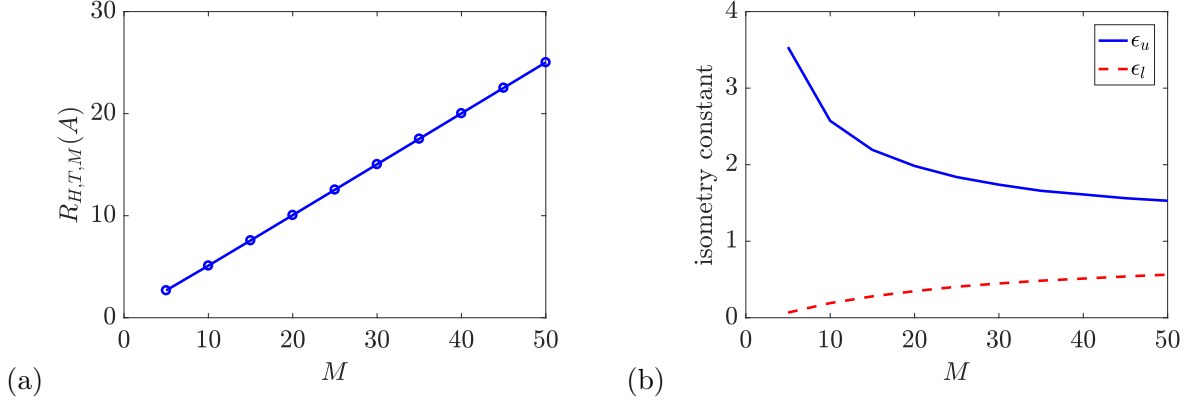
FIG. 2. *Stable rank and quality of delay-coordinate mapping for the linear system described in Section IV A. (a) Stable rank versus M (number of delays) with sampling interval $T \approx 1/250$. Note that the stable rank of the system gradually improves with increasing M. (b) Quality of embedding through delay-coordinate mapping as measured by the isometry constants $\epsilon_l \leq \epsilon_u$ versus M with $T \approx 1/250$ (see (2)). Note that, like the stable rank, the quality of embedding gradually improves with increasing M.*

B. Nonlinear Schrödinger System

As a case study involving a nonlinear system, we consider a sequence of points on a trajectory generated by a certain partial differential equation, the Nonlinear Schrödinger (NLS) equation:

$$\mathrm{i}u_t(z,t) + \frac{1}{2}u_{zz}(z,t) + |u(z,t)|^2 u(z,t) = 0.$$

Here, t denotes the continuous time variable and z denotes the continuous space variable; u_t denotes the partial derivative of u with respect to t and u_{zz} denotes the second order partial derivative of u with respect to z; and we adopt the boundary conditions $u \to 0$ as $z \to \pm\infty$. Adapting the construction provided in [69, Chapter 19], we sample $N = 800$ points between $z = -30$ and $z = 30$ at each time to generate data in $\mathbb{C}^N$. Data is generated with a time step of 0.02 seconds. The evolution of the trajectory over time is shown in Figure 3, which plots the magnitude of the entries of each data vector. The three different plots correspond to three different integer values of a parameter S which is used in the initial conditions

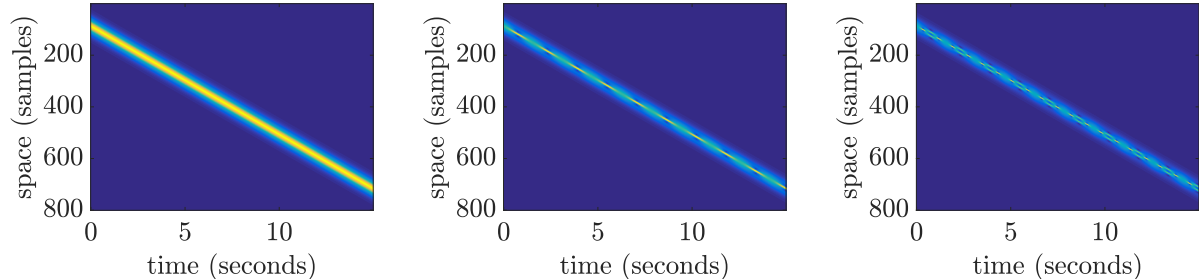$$u(z,0) = S\mathrm{sech}(z+z_0)e^{\mathrm{i}\Omega t}.$$

28

FIG. 3. *Magnitude of trajectory vectors for the Nonlinear Schrödinger system, with (from left to right) $S = 1$, $S = 2$, and $S = 3$-soliton solutions displayed.*

The resulting solutions are known as the S-soliton solutions (with $S = 1, 2, 3$) and have an initial center position $z_0 = 23.5$ and a drift over time due to the group velocity parameter $\Omega = \pi$.

We begin by presenting a specific example involving the $S = 2$-soliton solution. Figure 4(a) plots a random projection of the data vectors from $\mathbb{C}^N$ to $\mathbb{R}^3$, and Figure 4(b) shows the pairwise distances $\|x - y\|_2^2$ between all pairs $x, y \in \mathbb{C}^N$ on the trajectory. This projection, which is useful for obtaining a generic visualization of the trajectory, is computed by constructing a real-valued $3 \times N$ matrix Ξ populated with independent zero-mean and unit-variance Gaussian random variables. For each data vector $x \in \mathbb{C}^N$, we compute Ξx and preserve the real part of the resulting vector. Here, we consider only the final ~ 5 seconds of the data; the initial ~ 10 seconds are used for populating delay coordinate vectors when needed.

In this example, we set $T = 0.06$ seconds and $M = 160$. To construct the class $\mathcal{H}$ of measurement functions, we consider a set of $P = 50$ nonlinear *radial basis functions (RBFs)*, each defined by a center $v_p \in \mathbb{C}^N$ randomly chosen from a ball with radius comparable to the data set. The resulting measurement basis functions take the form $h_p(x) = e^{-\|x-v_p\|^2/2\sigma^2}$, where σ is a scaling parameter chosen to be comparable to the norm of a typical data vector.

Figure 4(c) shows a random projection of the resulting delay coordinate vectors $F_{h_\alpha, T, M}(x)$, where the entries of α are independent Gaussian random variables with zero mean and unit variance. Figure 4(d) shows the pairwise distances $\|F_{h_\alpha, T, M}(x) - F_{h_\alpha, T, M}(y)\|_2^2$ between all points in the reconstruction space. Figure 5(a) shows a scatter plot comparing the original distances $\|x - y\|_2^2$ between points on the trajectory to the corresponding distances $\|F_{h_\alpha, T, M}(x) - F_{h_\alpha, T, M}(y)\|_2^2$ in the reconstruction space $\mathbb{R}^M$. The
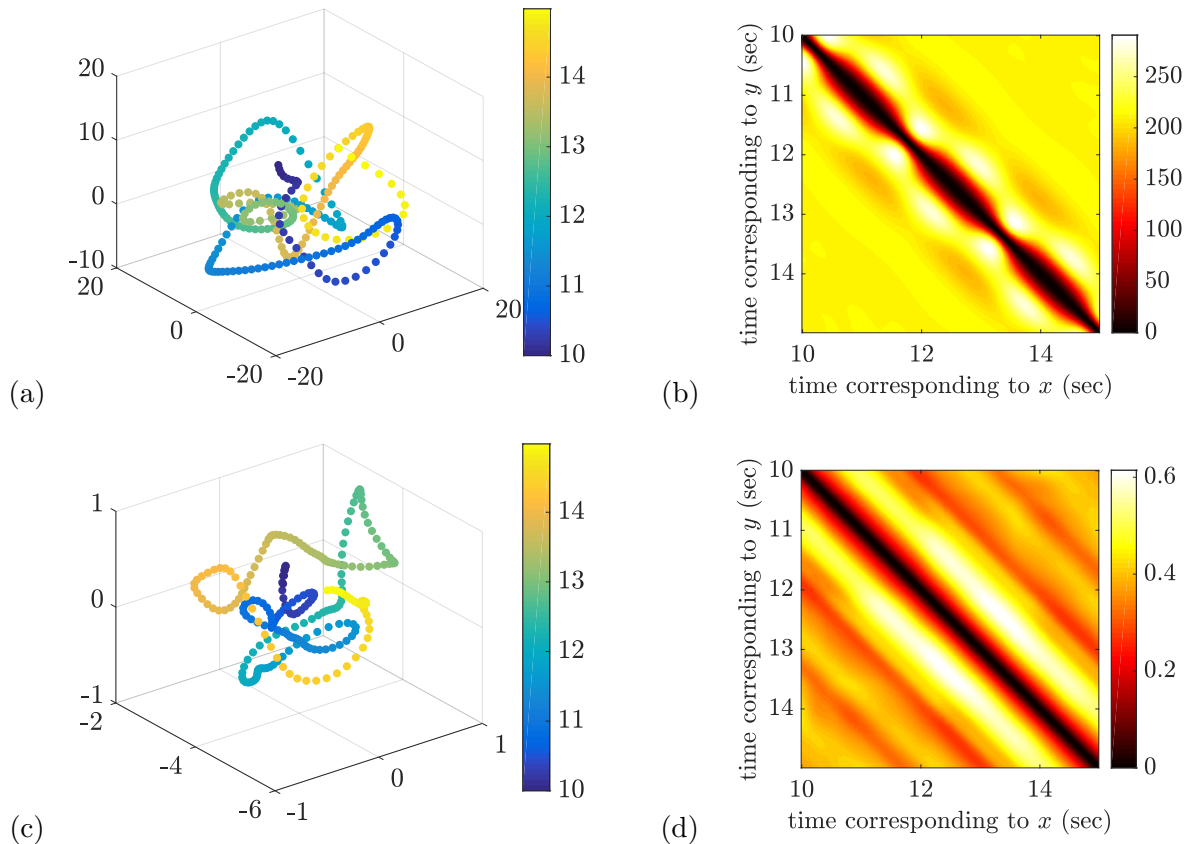
29

FIG. 4. *Embedding demonstration for the Nonlinear Schrödinger system, $S = 2$-soliton solution with nonlinear RBF measurement functions. (a) Visualization of data vectors $x \in \mathbb{C}^N$ on the original trajectory, projected via a random linear map to $\mathbb{R}^3$. Color (shading) is used to indicate the time (in seconds) corresponding to each data vector x. (b) Pairwise distances $\|x - y\|_2^2$ between all pairs $x, y \in \mathbb{C}^N$ on the original trajectory. (c) Visualization of the resulting delay coordinate vectors $F_{h_\alpha,T,M}(x)$, projected via a random linear map to $\mathbb{R}^3$. Again, color (shading) is used to indicate the time (in seconds) corresponding to each original data vector x. (d) Pairwise distances $\|F_{h_\alpha,T,M}(x) - F_{h_\alpha,T,M}(y)\|_2^2$ between all points in the reconstruction space.*

dashed lines have slopes equal to the minimum and maximum observed values of the ratio $\|F_{h_\alpha,T,M}(x) - F_{h_\alpha,T,M}(y)\|_2^2 / \|x - y\|_2^2$ over all pairs $x, y \in \mathbb{C}^N$ on the trajectory. Under a highly stable embedding (and in particular if the left and right hand sides of (20) were comparable to one another), the two lines in Figure 5(a) would have slopes comparable to one another. In this experiment, the ratio of the larger slope to the smaller slope is approximately 8.80. Up to some degree of approximation, small pairwise distances remain
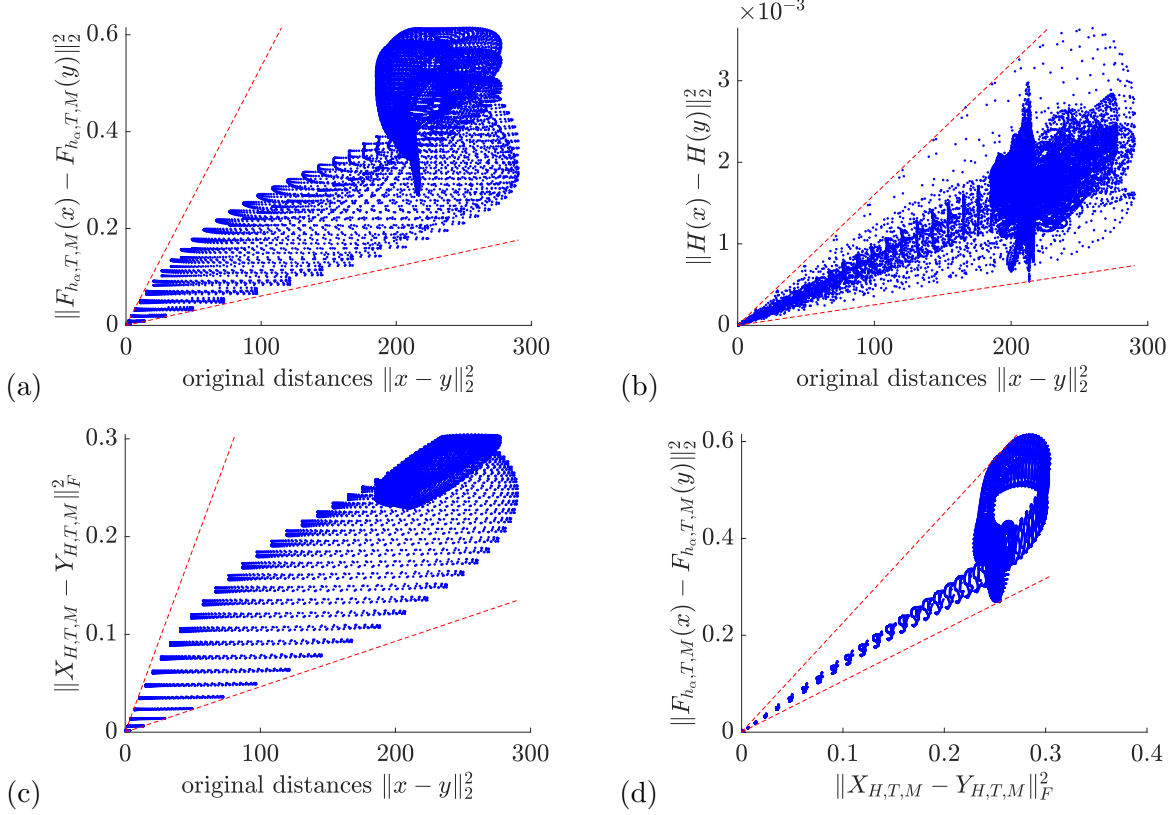
FIG. 5. *Pairwise distance preservation in the various stages of a delay coordinate embedding, $S = 2$-soliton solution with nonlinear RBF measurement functions.*

small, and large pairwise distances remain large.

We can unpack the factors that affect the degree of tightness in this embedding. The careful reader of Appendix A and especially (A29) will note that the variability of the ratio $\|F_{h_\alpha,T,M}(x) - F_{h_\alpha,T,M}(y)\|_2^2 / \|x - y\|_2^2$ is affected, in turn, by the variability of the ratios $\|H(x) - H(y)\|_2^2 / \|x - y\|_2^2$ (see assumption A.1 in Section III A as well as (A27)), $\|X_{H,T,M} - Y_{H,T,M}\|_F^2 / \|x - y\|_2^2$ (see (A28)), and $\|F_{h_\alpha,T,M}(x) - F_{h_\alpha,T,M}(y)\|_2^2 / \|X_{H,T,M} - Y_{H,T,M}\|_F^2$ (see (A23)). Figures 5(b), (c), and (d) show scatter plots corresponding to these three sets of pairwise distances, respectively. Variability in $\|H(x) - H(y)\|_2^2 / \|x - y\|_2^2$ can be caused by a large ratio between u_H and l_H; the ratio of the slopes in Figure 5(b) is approximately 6.35. Variability in $\|X_{H,T,M} - Y_{H,T,M}\|_F^2 / \|x - y\|_2^2$ is affected not only by u_H and l_H, but also by $\sigma_{\min}$, $\sigma_{\max}$, and geo($\mathbb{A}$). The ratio of the slopes in Figure 5(c) is approximately 8.05. Variability in $\|F_{h_\alpha,T,M}(x) - F_{h_\alpha,T,M}(y)\|_2^2 / \|X_{H,T,M} - Y_{H,T,M}\|_F^2$ is affected by the stable rank $\mathrm{R}_{H,T,M}(\mathbb{A})$, which in this example is approximately 5.66. The ratio of the slopes in Figure 5(d) is approximately 2.14. The tests below further reveal the causes and effects of

31

changing the parameters we have discussed here.

To further our study, we also experiment with the $S = 1$-soliton and $S = 3$-soliton solutions, and we test additional classes $\mathcal{H}$ of measurement functions. In addition to the RBF kernel already considered, we also take $\mathcal{H}$ to be the space of all linear functionals on $\mathbb{C}^N$, as in Section IV A. We also consider a set of nonlinear monomials of maximum degree K in N variables. Any such monomial can be written as $h_p(x) = x[1]^{\beta_1} \times x[2]^{\beta_2} \times \cdots \times x[N]^{\beta_N}$ for some $\{\beta_n\}_n$ with $\sum_n \beta_n \leq K$. We use a set of $P = 200$ randomly-picked monomials with maximum degree $K = 3$.

With a fixed value of $T = 0.06$ seconds, Figure 6(a) plots, as a function of M, the ratio of the largest and smallest isometry constants corresponding to $\|X_{H,T,M} - Y_{H,T,M}\|_F^2 / \|x - y\|_2^2$. (In the previous example, this corresponded to the ratio of the slopes in Figure 5(c), which was approximately 8.05.) Figure 6(b) shows the corresponding plot for $\|F_{h_\alpha,T,M}(x) - F_{h_\alpha,T,M}(y)\|_2^2 / \|X_{H,T,M} - Y_{H,T,M}\|_F^2$, and Figure 6(c) shows the corresponding plot for $\|F_{h_\alpha,T,M}(x) - F_{h_\alpha,T,M}(y)\|_2^2 / \|x - y\|_2^2$, reflecting the tightness of the overall embedding. Figure 6(d) shows the stable rank $\mathrm{R}_{H,T,M}(\mathbb{A})$ as a function of M. In these plots, we see several general trends:

- The overall embedding is generally tightest for the $S = 1$-soliton solution and weakest for the $S = 3$-soliton solution. As illustrated in Figure 3, the complexity of the trajectories generally increases for larger values of S. For example, the trajectory for $S = 1$ has constant speed, while the instantaneous speed of the trajectory when $S = 3$ varies over a dynamic range of approximately 6.95. This variability affects factors such as $\sigma_{\min}$ and $\sigma_{\max}$, leading to more variability in $\|X_{H,T,M} - Y_{H,T,M}\|_F^2 / \|x - y\|_2^2$ as shown in Figure 6(a). There is relatively little effect of S on $\|F_{h_\alpha,T,M}(x) - F_{h_\alpha,T,M}(y)\|_2^2 / \|X_{H,T,M} - Y_{H,T,M}\|_F^2$ as shown in Figure 6(b).

- The linear measurement functions generally result in the tightest embeddings; partly this is due to the fact that $l_H = u_H = 1$ in the linear case. The nonlinear monomial functions produce the loosest embeddings. However, the nonlinear RBF functions perform nearly as well as the linear functions.

- In general, as M increases, the stable rank increases $\mathrm{R}_{H,T,M}(\mathbb{A})$, which reduces the variability of $\|F_{h_\alpha,T,M}(x) - F_{h_\alpha,T,M}(y)\|_2^2 / \|X_{H,T,M} - Y_{H,T,M}\|_F^2$ and thus of the overall
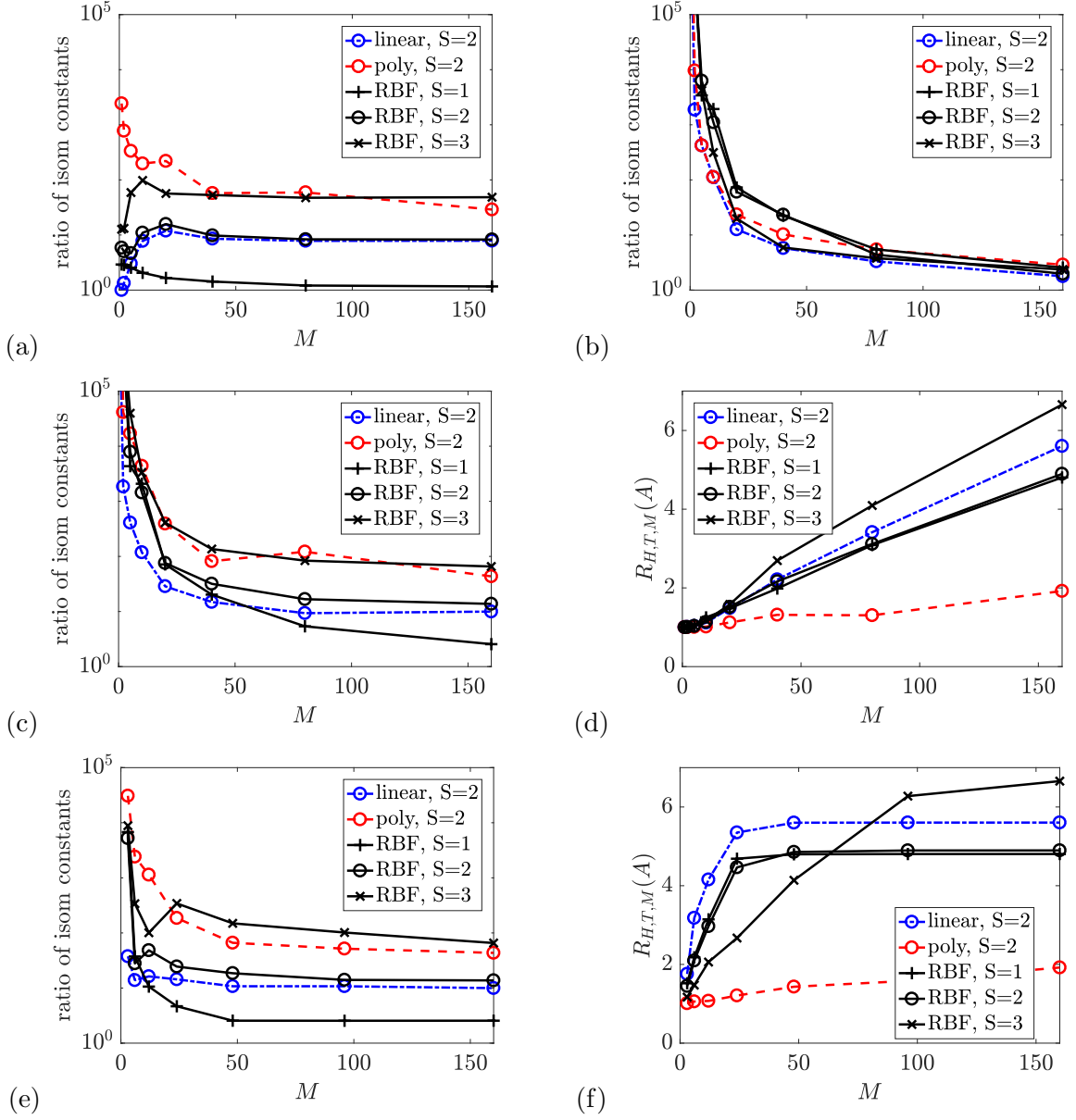
FIG. 6. *Embedding demonstration for the Nonlinear Schrödinger system, including the $S = 1$, $S = 2$, and $S = 3$-soliton solutions and various linear and nonlinear measurement functions. With T fixed to 0.06 seconds: (a) $x - y$ vs. $X_{H,T,M} - Y_{H,T,M}$, (b) $X_{H,T,M} - Y_{H,T,M}$ vs. $F_{h_\alpha,T,M}(x) - F_{h_\alpha,T,M}(y)$, (c) $x - y$ vs. $F_{h_\alpha,T,M}(x) - F_{h_\alpha,T,M}(y)$, (d) stable rank. With MT fixed to 9.6 seconds: (e) $x - y$ vs. $F_{h_\alpha,T,M}(x) - F_{h_\alpha,T,M}(y)$, (f) stable rank.*

embedding $\|F_{h_\alpha,T,M}(x) - F_{h_\alpha,T,M}(y)\|_2^2 / \|x - y\|_2^2$. This is as expected in light of Theorem III.1.

Finally, over a fixed total delay of 9.6 seconds, we experiment with a range of M values.

33

In each case, we choose $T = 9.6/M$. Results are shown in Figure 6(e),(f). These results show that, over this total amount of time it is not necessary to sample densely in time; moderately small values of M (around 40), corresponding to moderately large values of T (around 0.24 seconds) lead to delay coordinate embeddings with a reasonable degree of tightness.

V. CONCLUSIONS AND OPEN PROBLEMS

The main result of this paper extends Takens' celebrated embedding theorem to provide conditions when a delay-coordinate map can provide a stable embedding of a dynamical system attractor. Given the prevalence of these techniques in nonlinear time-series analysis, this result provides a much needed theoretical justification of their numerical performance in the presence of real-world imperfections such as noise and quantization. While the conditions of this result are restrictive and it may not always be possible to meet them in practice, there is significant value in knowing for which scenarios one can guarantee a given quality level of the embedding. In fact, researchers have informally conjectured that instability issues may limit the performance of numerical techniques based on delay-coordinate mapping without the theoretical foundations to examine this issue formally (e.g., see the discussion regarding Takens' theorem in the Supporting Online Material of [70]). The examination of our results has also led to new and insightful interpretations of classical (generally heuristic) techniques for selecting parameters such as the sampling time and number of delays.

Building on these results, there appears to be no shortage of interesting directions for future work. For example:

- Remark III.3 and Table I provide a recipe for choosing the sampling interval T and the number of delays M in delay-coordinate mapping. It is of interest to experimentally validate this procedure and perhaps find alternatives with lower computational complexity.

- An open question is whether it is possible to improve (increase) the stable rank of a dynamical system (and hence improve the quality of delay-coordinate mapping) by optimizing over the choice of scalar measurement functions. While we suspect that answer is negative, a rigorous study of this topic does not currently exist.

- A remaining technical challenge is the role of $\text{rch}(\mathbb{A}_{H,T,M})$ (reach of the "trajectory

attractor") in Theorem III.1 (also see (13)). We suspect that $\text{rch}(\mathbb{A}_{H,T,M})$ can be expressed entirely in terms of $\text{rch}(\mathbb{A})$ (and T, M, and basis functions $\{h_p\}$). Such an expression will substantially simplify and clarify Theorem III.1 but has remained elusive despite our efforts.

- While multivariate time-series have been occasionally discussed in the literature (e.g., [71]), as with our work, most treatments of delay-coordinate maps are restricted to a single scalar measurement function. An open question is how the presence of multiple measurement functions (producing diverse observations at each sampling time step) would affect the stability of the attractor embedding.

Appendix A: Proof of Theorem III.1 (Stable Takens' Theorem)

We reserve the letters $C, C_1, C_2, \cdots$ to represent universal positive constants. We adopt the following (semi-)order: $a \lesssim b$ means that there is an absolute constant C_1 such that $a \leq C_1 b$. If, instead of being an absolute constant, $C_1 = C_1(\theta)$ depends on some parameter θ, we write $a \lesssim_\theta b$. Of course, $a \gtrsim b$ and $a \gtrsim_\theta b$ are defined similarly. Occasionally, we will use the convention that $[a : b] = a, a + 1, \cdots, b$ for integers $a \leq b$.

Throughout the proof, the dependence on different quantities might be suppressed if there is no ambiguity. Consider $x \in \mathbb{A}$ and scalar measurement function $h_\alpha(\cdot) = \sum_p \alpha[p] h_p(\cdot)$ as a linear combination of basis functions. Recall from (11) and (12) that the corresponding delay vector can be written as

$$F_{h_\alpha, T, M}(x) = X_{H,T,M}\alpha,$$

35

$$X_{H,T,M} = \left[H(x) \; H\left(\phi_T^{-1}(x)\right) \; \cdots \; H\left(\phi_T^{-M+1}(x)\right) \right]^* \in \mathbb{R}^{M \times P}, \tag{A1}$$

and where $\phi_T : \mathbb{A} \to \mathbb{A}$ is the flow on the attractor. For a fixed pair of points $x, y \in \mathbb{A}$, consider the random variable

$$\frac{\|F_{h_\alpha,T,M}(x) - F_{h_\alpha,T,M}(y)\|_2^2}{\|X_{H,T,M} - Y_{H,T,M}\|_F^2} = \frac{\|X_{H,T,M}\alpha - Y_{H,T,M}\alpha\|_2^2}{\|X_{H,T,M} - Y_{H,T,M}\|_F^2}, \tag{A2}$$

and note that

$$\mathbb{E}\left[\frac{\|F_{h_\alpha,T,M}(x) - F_{h_\alpha,T,M}(y)\|_2^2}{\|X_{H,T,M} - Y_{H,T,M}\|_F^2}\right] = \frac{\mathbb{E}\left[\|(X_{H,T,M} - Y_{H,T,M})\alpha\|_2^2\right]}{\|X_{H,T,M} - Y_{H,T,M}\|_F^2} = 1,$$

where the second identity holds because the entries of $\alpha \in \mathbb{R}^P$ are independent and have unit-variance. This suggests that for *all* pairs of points in $\mathbb{A}$, the ratio in (A2) might be close to one. That is, we hope that the following quantity is small with overwhelming probability:

$$\sup_{x,y \in \mathbb{A}} \left| \frac{\|F_{h_\alpha,T,M}(x) - F_{h_\alpha,T,M}(y)\|_2^2}{\|X_{H,T,M} - Y_{H,T,M}\|_F^2} - 1 \right| = \sup_{Z \in \mathfrak{Z}} \left| \|Z\alpha\|_2^2 - \mathbb{E}\left[\|Z\alpha\|_2^2\right] \right|. \tag{A3}$$

Above, we conveniently set

$$\mathfrak{Z} := \left\{ \frac{X_{H,T,M} - Y_{H,T,M}}{\|X_{H,T,M} - Y_{H,T,M}\|_F} : x, y \in \mathbb{A} \right\} \subset \mathbb{R}^{M \times P}. \tag{A4}$$

To control the supremum in (A3), we invoke a recent result by Krahmer et al.

Proposition A.1. [55, Theorem 3.1] For integers M and P, let $\mathfrak{Z} \subset \mathbb{R}^{M \times P}$ be a collection of matrices. Moreover, let $\alpha \in \mathbb{R}^P$ be a random vector whose entries are independent zero-mean, unit-variance random variables with subgaussian norm of θ. Set

$$d_F(\mathfrak{Z}) = \sup_{Z \in \mathfrak{Z}} \|Z\|_F, \qquad d_2(\mathfrak{Z}) = \sup_{Z \in \mathfrak{Z}} \|Z\|,$$

where $\| \cdot \|_F$ and $\| \cdot \|$ stand for the Frobenius and spectral norms, respectively. Also let $\gamma_2(\mathfrak{Z}, \| \cdot \|)$ be the Gaussian width of $\mathfrak{Z}$ with respect to the spectral metric, and define

$$E_1 := \gamma_2\left(\mathfrak{Z}, \| \cdot \|\right) \cdot \left(\gamma_2\left(\mathfrak{Z}, \| \cdot \|\right) + d_F(\mathfrak{Z})\right) + d_F(\mathfrak{Z})d_2(\mathfrak{Z}),$$

$$E_2 := d_2^2(\mathfrak{Z}),$$

$$E_3 := d_2(\mathfrak{Z})\left(\gamma_2\left(\mathfrak{Z}, \| \cdot \|\right) + d_F(\mathfrak{Z})\right).$$

Then, for arbitrary $v > 0$, it holds that

$$\mathbb{P}\left[\sup_{Z \in \mathfrak{Z}} \left| \|Z\alpha\|_2^2 - \mathbb{E}\left[\|Z\alpha\|_2^2\right] \right| > C_2(\theta)E_1 + v\right] \leq 2\exp\left(-C_3(\theta)\min\left[\frac{v}{E_2}, \frac{v^2}{E_3^2}\right]\right),$$

where $C_2(\theta)$ and $C_3(\theta)$ depend only on θ.

Without dwelling too much on the concept of Gaussian width above, we recall the following well-known relation [72]:

$$\gamma_2\left(\mathfrak{Z}, \|\cdot\|\right) \lesssim \int_0^\infty \sqrt{\log\left(\#\left(\mathfrak{Z}, \|\cdot\|, s\right)\right)}\, ds. \tag{A5}$$

Above, $\#(\mathfrak{Z}, \|\cdot\|, s)$ is the *covering number* of the set $\mathfrak{Z}$ with respect to the metric $\|\cdot\|$ and at scale $s > 0$. That is, $\#(\mathfrak{Z}, \|\cdot\|, s)$ is the smallest number of balls of radius s (and with respect to the metric $\|\cdot\|$) needed to cover the set $\mathfrak{Z}$. In order to apply Proposition A.1 to (A3), we must first calculate $d_F(\mathfrak{Z})$, $d_2(\mathfrak{Z})$, and $\#(\mathfrak{Z}, \|\cdot\|, s)$ (with $\mathfrak{Z}$ as in (A4)). That, we set out to do next. Observe that

$$d_F\left(\mathfrak{Z}\right) = \sup_{Z\in\mathfrak{Z}}\|Z\|_F = \sup_{x,y\in\mathbb{A}} \frac{\|X_{H,T,M} - Y_{H,T,M}\|_F}{\|X_{H,T,M} - Y_{H,T,M}\|_F} = 1, \qquad \text{(see (A4))} \tag{A6}$$

$$d_2\left(\mathfrak{Z}\right) = \sup_{Z\in\mathfrak{Z}}\|Z\| = \sup_{x,y\in\mathbb{A}} \frac{\|X_{H,T,M} - Y_{H,T,M}\|}{\|X_{H,T,M} - Y_{H,T,M}\|_F} = \frac{1}{\sqrt{\mathrm{R}_{H,T,M}\left(\mathbb{A}\right)}}. \qquad \text{(see (16))} \tag{A7}$$

Estimating the covering number of $\mathfrak{Z}$ is more involved. From the order between norms $\|\cdot\| \leq \|\cdot\|_F$, first deduce that

$$\#\left(\mathfrak{Z}, \|\cdot\|, s\right) \leq \#\left(\mathfrak{Z}, \|\cdot\|_F, s\right), \qquad \forall s > 0. \tag{A8}$$

The covering number on the right hand side above is easier to control, as $\mathfrak{Z} \subset \mathbb{R}^{M\times P}$ is isometric to another (more malleable) object that we denote here with $U(\mathbb{A}_{H,T,M})$ and define next. Set

$$x_{H,T,M} = x_{H,T,M}(x) := \mathrm{vec}\left(X_{H,T,M}\right) = \begin{bmatrix} H(x) \\ \vdots \\ H\left(\phi_T^{-M+1}(x)\right) \end{bmatrix} \in \mathbb{R}^{MP}, \qquad \forall x \in \mathbb{A}, \tag{A9}$$

$$\mathbb{A}_{H,T,M} := \left\{x_{H,T,M} : x \in \mathbb{A}\right\} \subset \mathbb{R}^{MP}. \tag{A10}$$

Then, let $U(\mathbb{A}_{H,T,M})$ denote the set of all directions in $\mathbb{A}_{H,T,M}$, i.e.,

$$U\left(\mathbb{A}_{H,T,M}\right) := \left\{\frac{x_{H,T,M} - y_{H,T,M}}{\|x_{H,T,M} - y_{H,T,M}\|_2} : x, y \in \mathbb{A}\right\} \subset \S^{MP-1}, \tag{A11}$$

where $\S^{MP-1}$ is the unit sphere in $\mathbb{R}^{MP}$. Recalling (A4), we observe that the pair $(\mathfrak{Z}, \|\cdot\|_F)$ is isometric to the pair $(U(\mathbb{A}_{H,T,M}), \|\cdot\|_2)$. Thanks to this isometry, we may continue to simplify (A8) by writing that

$$\#\left(\mathfrak{Z}, \|\cdot\|, s\right) \leq \#\left(\mathfrak{Z}, \|\cdot\|_F, s\right)$$
$$= \#\left(U\left(\mathbb{A}_{H,T,M}\right), \|\cdot\|_2, s\right). \tag{A12}$$

37

Next, we estimate the covering number of $U(\mathbb{A}_{H,T,M})$. Recall that the attractor $\mathbb{A} \subset \mathbb{R}^N$ is a well-behaved manifold and the flow $\phi_T(\cdot)$ is a diffeomorphism on $\mathbb{A}$. Not surprisingly, then, $\mathbb{A}_{H,T,M}$ (defined in (A10)) too is a well-behaved manifold whose geometrical attributes can be expressed in terms of those of $\mathbb{A}$. This observation is formalized next and proved in Appendix B.

Lemma A.1. *Recall the attractor* $\mathbb{A} \subset \mathbb{R}^N$, *and the flow* $\phi_T : \mathbb{A} \to \mathbb{A}$, *which by assumption is a diffeomorphism on* $\mathbb{A}$. *Let* $D\phi_T(x) : \mathbb{T}_x\mathbb{A} \to \mathbb{T}_{\phi_T(x)}\mathbb{A}$ *be the derivative of the flow at* $x \in \mathbb{A}$ *(see Section II C). The linear map* $D\phi_T(x)$ *may be identified with a* $\dim(\mathbb{A}) \times \dim(\mathbb{A})$ *matrix. Assume that the singular values of this matrix belong to some interval* $[\sigma_{\min}, \sigma_{\max}] \subset (0, \infty)$. *Lastly, recall the properties of the map* $H(\cdot)$ *listed in Section III A.*

Then, $\mathbb{A}_{H,T,M} \subset \mathbb{R}^{MP}$, *as specified in (A10), is a bounded, boundary-less, and smooth submanifold of* $\mathbb{R}^{MP}$ *with* $\dim(\mathbb{A}_{H,T,M}) = \dim(\mathbb{A})$. *Moreover,*

$$\eta_{\min}^{\dim(\mathbb{A})} \sqrt{\frac{\sigma_{\max}^{-2M\dim(\mathbb{A})} - 1}{\sigma_{\max}^{-2\dim(\mathbb{A})} - 1}} \, \mathrm{vol}\,(\mathbb{A}) \leq \mathrm{vol}\,(\mathbb{A}_{H,T,M}) \leq \eta_{\max}^{\dim(\mathbb{A})} \sqrt{\frac{\sigma_{\min}^{-2M\dim(\mathbb{A})} - 1}{\sigma_{\min}^{-2\dim(\mathbb{A})} - 1}} \, \mathrm{vol}\,(\mathbb{A}).$$

The above lemma controls the geometric properties of $U(\mathbb{A}_{H,T,M})$—its dimension and volume. By substituting these estimates into Lemma 15 of [33], we can in turn control the covering number of $U(\mathbb{A}_{H,T,M})$ by writing that

$$\# \left(U\left(\mathbb{A}_{H,T,M}\right), \|\cdot\|_2, s \right)$$

$$\leq 2 \left(\frac{6.12\sqrt{\dim\left(\mathbb{A}_{H,T,M}\right)}}{s^2} \right)^{2\dim\left(\mathbb{A}_{H,T,M}\right)}$$

$$\left(\frac{\mathrm{vol}\left(\mathbb{A}_{H,T,M}\right)}{\mathrm{rch}\left(\mathbb{A}_{H,T,M}\right)^{\dim(\mathbb{A}_{H,T,M})}} \right)^2 \qquad (\text{invoke } [33, \text{Lemma } 15])$$

$$\leq 2 \left(\frac{6.12\sqrt{\dim\left(\mathbb{A}\right)}}{s^2} \right)^{2\dim(\mathbb{A})} \eta_{\max}^{2\dim(\mathbb{A})}$$

$$\frac{\sigma_{\min}^{-2M\dim(\mathbb{A})} - 1}{\sigma_{\min}^{-2\dim(\mathbb{A})} - 1} \left(\frac{\mathrm{vol}\left(\mathbb{A}\right)}{\mathrm{rch}\left(\mathbb{A}_{H,T,M}\right)^{\dim(\mathbb{A})}} \right)^2, \quad (\text{invoke Lemma A.1})$$

which holds for any $s \leq \frac{1}{2}$, and under the mild assumption that the volume of $\mathbb{A}$ is sufficiently

large:

$$\frac{\mathrm{vol}\left(\mathbb{A}_{H,T,M}\right)}{\mathrm{rch}\left(\mathbb{A}_{H,T,M}\right)^{\dim\left(\mathbb{A}_{H,T,M}\right)}} \geq \eta_{\min}^{\dim(\mathbb{A})} \sqrt{\frac{\sigma_{\max}^{-2M\dim(\mathbb{A})}-1}{\sigma_{\max}^{-2\dim(\mathbb{A})}-1}} \frac{\mathrm{vol}\left(\mathbb{A}\right)}{\mathrm{rch}\left(\mathbb{A}_{H,T,M}\right)^{\dim(\mathbb{A})}}$$

$$\geq \left(\frac{21}{2\sqrt{\dim\left(\mathbb{A}_{H,T,M}\right)}}\right)^{\dim\left(\mathbb{A}_{H,T,M}\right)}$$

$$= \left(\frac{21}{2\sqrt{\dim\left(\mathbb{A}\right)}}\right)^{\dim(\mathbb{A})}.$$

In light of (A12), we conclude that

$$\#\left(\mathfrak{Z},\|\cdot\|,s\right) \leq 2\left(\frac{6.12\sqrt{\dim\left(\mathbb{A}\right)}}{s^2}\right)^{2\dim(\mathbb{A})} \eta_{\max}^{2\dim(\mathbb{A})} \frac{\sigma_{\min}^{-2M\dim(\mathbb{A})}-1}{\sigma_{\min}^{-2\dim(\mathbb{A})}-1} \left(\frac{\mathrm{vol}\left(\mathbb{A}\right)}{\mathrm{rch}\left(\mathbb{A}_{H,T,M}\right)^{\dim(\mathbb{A})}}\right)^2,$$

and we denote the right hand side by

$$\left(\frac{\Delta}{s}\right)^{4\dim(\mathbb{A})}. \tag{A13}$$

The above bound holds for every $s \leq \frac{1}{2}$, and as long as

$$\frac{\mathrm{vol}\left(\mathbb{A}\right)^{\frac{1}{\dim(\mathbb{A})}}}{\mathrm{rch}\left(\mathbb{A}_{H,T,M}\right)} \geq \eta_{\min}^{-1}\left(\frac{\sigma_{\max}^{-2M\dim(\mathbb{A})}-1}{\sigma_{\max}^{-2\dim(\mathbb{A})}-1}\right)^{-\frac{1}{2\dim(\mathbb{A})}} \frac{21}{2\sqrt{\dim\left(\mathbb{A}\right)}}. \tag{A14}$$

With the covering number of $\mathfrak{Z}$ at hand, we now use (A5) to control the Gaussian width of

39

$\mathfrak{Z}$:

$$\gamma_2\left(\mathfrak{Z}, \|\cdot\|\right)$$

$$\lesssim \int_0^\infty \sqrt{\log\left(\#\left(\mathfrak{Z}, \|\cdot\|, s\right)\right)}\, ds$$

$$= \int_0^{2d_2(\mathfrak{Z})} \sqrt{\log\left(\#\left(\mathfrak{Z}, \|\cdot\|, s\right)\right)}\, ds \qquad \left(\text{from (A7): } s \geq 2d_2\left(\mathfrak{Z}\right) \implies \#\left(\mathfrak{Z}, \|\cdot\|, s\right) = 1\right)$$

$$\leq \int_0^{\frac{2}{\sqrt{\mathrm{R}_{H,T,M}(\mathbb{A})}}} \sqrt{\log\left(\#\left(\mathfrak{Z}, \|\cdot\|, s\right)\right)}\, ds \qquad (\text{see (A7)})$$

$$\leq \sqrt{4\dim(\mathbb{A})} \int_0^{\frac{2}{\sqrt{\mathrm{R}_{H,T,M}(\mathbb{A})}}} \sqrt{\log\left(\frac{\Delta}{s}\right)}\, ds \qquad (\text{see (A13)})$$

$$\leq \sqrt{4\dim(\mathbb{A})} \left(\frac{4}{\mathrm{R}_{H,T,M}(\mathbb{A})}\right)^{\frac{1}{4}}$$

$$\sqrt{\int_0^{\frac{2}{\sqrt{\mathrm{R}_{H,T,M}(\mathbb{A})}}} \log\left(\frac{\Delta}{s}\right) ds} \qquad \left(\int_0^a \sqrt{f(s)}\, ds \leq \sqrt{a \int_0^a f(s)\, ds}\right)$$

$$\leq \sqrt{4\dim(\mathbb{A})} \left(\frac{4}{\mathrm{R}_{H,T,M}(\mathbb{A})}\right)^{\frac{1}{4}} \sqrt{\int_0^{\frac{2}{\sqrt{\mathrm{R}_{H,T,M}(\mathbb{A})}}} \log\left(1 + \frac{\Delta}{s}\right) ds}$$

$$\leq \sqrt{4\dim(\mathbb{A})} \left(\frac{4}{\mathrm{R}_{H,T,M}(\mathbb{A})}\right)^{\frac{1}{4}} \sqrt{\frac{4}{\sqrt{\mathrm{R}_{H,T,M}(\mathbb{A})}}}$$

$$\sqrt{\log\left(1 + \frac{\Delta\sqrt{\mathrm{R}_{H,T,M}(\mathbb{A})}}{2}\right)} \quad \left(\int_0^a \log\left(1 + \frac{b}{s}\right) ds \leq 2a \log\left(1 + \frac{b}{a}\right), \text{ if } a \leq b\right)$$

$$\leq 8\sqrt{\frac{\dim(\mathbb{A})}{\mathrm{R}_{H,T,M}(\mathbb{A})} \log\left(\Delta\sqrt{\mathrm{R}_{H,T,M}(\mathbb{A})}\right)} \qquad (\log(1+a) \leq 2\log(a), \forall a \geq 2) \qquad (\text{A15})$$

and, by simplifying the last line,

$$\gamma_2\left(\mathcal{Z}, \|\cdot\|\right)$$

$$\lesssim \sqrt{\frac{\dim(\mathbb{A})}{\mathrm{R}_{H,T,M}(\mathbb{A})}}$$

$$\sqrt{\log\left(\sqrt{\dim(\mathbb{A})}\eta_{\max}\left(\frac{\sigma_{\min}^{-2M\dim(\mathbb{A})} - 1}{\sigma_{\min}^{-2\dim(\mathbb{A})} - 1}\right)^{\frac{1}{2\dim(\mathbb{A})}} \frac{\mathrm{vol}(\mathbb{A})^{\frac{1}{\dim(\mathbb{A})}}}{\mathrm{rch}(\mathbb{A}_{H,T,M})} \mathrm{R}_{H,T,M}(\mathbb{A})\right)}. \,(\text{see (A13)})$$

$$(\text{A16})$$

For the fifth and tenth lines to hold, we must impose that

$$\mathrm{R}_{H,T,M}(\mathbb{A}) \geq 16\max\left(1, \Delta^{-2}\right). \qquad (\text{A17})$$

For (A17) to hold, it actually suffices to assume that

$$R_{H,T,M}(\mathbb{A}) \gtrsim 1, \tag{A18}$$

$$\sqrt{\dim(\mathbb{A})} \eta_{\max} \left(\frac{\sigma_{\min}^{-2M\dim(\mathbb{A})} - 1}{\sigma_{\min}^{-2\dim(\mathbb{A})} - 1} \right)^{\frac{1}{2\dim(\mathbb{A})}} \frac{\mathrm{vol}(\mathbb{A})^{\frac{1}{\dim(\mathbb{A})}}}{\mathrm{rch}(\mathbb{A}_{H,T,M})} \gtrsim 1. \qquad \text{(see (A13))} \tag{A19}$$

We note that (A19) is guaranteed to hold if (A14) (which appears in the theorem statement as (19)) holds. Given the estimates of $d_F(\mathfrak{Z})$, $d_2(\mathfrak{Z})$, and $\gamma_2(\mathfrak{Z}, \|\cdot\|)$ (see (A6-A7) and (A16)), we are now in position to apply Proposition A.1 to $\mathfrak{Z}$ (specified in (A4)). For $\delta, \rho \in (0,1)$, assume that

$$R_{H,T,M}(\mathbb{A}) \gtrsim \delta^{-2} \dim(\mathbb{A}) \log \left(\sqrt{\dim(\mathbb{A})} \eta_{\max} \left(\frac{\sigma_{\min}^{-2M\dim(\mathbb{A})} - 1}{\sigma_{\min}^{-2\dim(\mathbb{A})} - 1} \right)^{\frac{1}{2\dim(\mathbb{A})}} \frac{\mathrm{vol}(\mathbb{A})^{\frac{1}{\dim(\mathbb{A})}}}{\mathrm{rch}(\mathbb{A}_{H,T,M})} \right.$$
$$\left. R_{H,T,M}(\mathbb{A}) \right) + \delta^{-2} \log \left(\frac{1}{\rho} \right). \tag{A20}$$

Under this assumption, we obtain that

$$d_F(\mathfrak{Z}) = 1, \qquad \text{(see (A6))}$$

$$d_2(\mathfrak{Z}) = \frac{1}{\sqrt{R_{H,T,M}(\mathbb{A})}} \lesssim \delta \left(\log \left(\frac{1}{\rho} \right) \right)^{-\frac{1}{2}}, \qquad \text{(see (A7))}$$

$$\gamma_2(\mathfrak{Z}, \|\cdot\|) \lesssim \delta. \quad \text{(see (A16))}$$

Subsequently, the quantities E_1, E_2, and E_3 in Proposition A.1 may be bounded as

$$E_1 \lesssim \delta(\delta + 1) + \delta \left(\log \left(\frac{1}{\rho} \right) \right)^{-\frac{1}{2}} \lesssim \delta, \qquad (\rho < 1)$$

$$E_2 \lesssim \delta^2 \left(\log \left(\frac{1}{\rho} \right) \right)^{-1},$$

$$E_3 \lesssim \delta \left(\log \left(\frac{1}{\rho} \right) \right)^{-\frac{1}{2}} (\delta + 1) \lesssim \delta \left(\log \left(\frac{1}{\rho} \right) \right)^{-\frac{1}{2}}.$$

We now recall (A3), substitute the above quantities into Proposition A.1 with an arbitrary

41

$v > 0$, and finally find that

$$\mathbb{P}\left[\sup_{x,y\in\mathbb{A}}\left|\frac{\|F_{h_\alpha,T,M}(x)-F_{h_\alpha,T,M}(y)\|_2^2}{\|X_{H,T,M}-Y_{H,T,M}\|_F^2}-1\right|>C_2(\theta)\delta+v\right]$$

$$=\mathbb{P}\left[\sup_{Z\in\mathfrak{Z}}\left|\|Z\alpha\|_2^2-\mathbb{E}\left[\|Z\alpha\|_2^2\right]\right|>C_2(\theta)\delta+v\right]\qquad\text{(see (A3))}$$

$$\leq\mathbb{P}\left[\sup_{Z\in\mathfrak{Z}}\left|\|Z\alpha\|_2^2-\mathbb{E}\left[\|Z\alpha\|_2^2\right]\right|>C_2(\theta)E_1+v\right]\qquad\text{(see (A21))}$$

$$\leq2\exp\left(-C_3(\theta)\min\left[\frac{v}{E_2},\frac{v^2}{E_3^2}\right]\right)\qquad\text{(see Proposition A.1)}$$

$$\leq2\exp\left(-\frac{C_3(\theta)}{\delta^2}\log\left(\frac{1}{\rho}\right)\min\left(v^2,v\right)\right).\qquad\text{(see (A21))}\qquad\text{(A21)}$$

By assigning $v=\delta$ above, we conclude that

$$\mathbb{P}\left[\sup_{x,y\in\mathbb{A}}\left|\frac{\|F_{h_\alpha,T,M}(x)-F_{h_\alpha,T,M}(y)\|_2^2}{\|X_{H,T,M}-Y_{H,T,M}\|_F^2}-1\right|>C_4(\theta)\delta\right]$$

$$\leq2\exp\left(-\frac{C_3(\theta)}{\delta^2}\log\left(\frac{1}{\rho}\right)\min\left(\delta^2,\delta\right)\right)$$

$$\leq C_5(\theta)\rho,\qquad(\delta<1)\qquad\text{(A22)}$$

for $C_4(\theta)$ and $C_5(\theta)$ that depend only on θ. Equivalently, if we replace $\gtrsim$ in (A20) with $\gtrsim_\theta$, we can further simplify the above inequality to read

$$\mathbb{P}\left[\sup_{x,y\in\mathbb{A}}\left|\frac{\|F_{h_\alpha,T,M}(x)-F_{h_\alpha,T,M}(y)\|_2^2}{\|X_{H,T,M}-Y_{H,T,M}\|_F^2}-1\right|>\delta\right]\leq C_6(\theta)\rho,\qquad\text{(A23)}$$

for $C_6(\theta)$ that depend only on θ. Here, $\gtrsim_\theta$ hides the explicit dependence on θ for convenience. This proves the version of Theorem III.1 that appears in Remark A.1.

Fix $x,y\in\mathbb{A}$. We can in fact replace $\|X_{H,T,M}-Y_{H,T,M}\|_F$ above with a more approachable quantity as follows. From (A1), recall that

$$\|X_{H,T,M}-Y_{H,T,M}\|_F^2=\sum_{m=0}^{M-1}\left\|H\left(\phi_T^{-m}(x)\right)-H\left(\phi_T^{-m}(y)\right)\right\|_2^2,\qquad\text{(A24)}$$

which suggests that we should find a more convenient expression for each summand above. Invoking the assumption in Theorem III.1 that the spectrum of $D\phi_T(\cdot)$ (the derivative of the flow) belongs to some interval $[\sigma_{\min},\sigma_{\max}]\in(0,\infty)$, we may easily verify that

$$\sigma_{\max}^{-m}d_\mathbb{A}\left(x,y\right)\leq d_\mathbb{A}\left(\phi_T^{-m}(x),\phi_T^{-m}(y)\right)\leq\sigma_{\min}^{-m}d_\mathbb{A}\left(x,y\right),\qquad\text{(A25)}$$

42

where $d_{\mathbb{A}}(\cdot, \cdot)$ returns the geodesic distance between a pair of points on $\mathbb{A}$ (see Section II C). To relate the geodesic metric on $\mathbb{A}$ to the Euclidean metric in $\mathbb{R}^N$, we recall the regularity condition (5), from which it follows that

$$\text{geo}(\mathbb{A})^{-1}\sigma_{\max}^{-m}\|x - y\|_2 \leq \left\|\phi_T^{-m}(x) - \phi_T^{-m}(y)\right\|_2 \leq \text{geo}(\mathbb{A})\sigma_{\min}^{-m}\|x - y\|_2. \tag{A26}$$

Next, recalling the bi-Lipschitz property of $H(\cdot)$ in Section III A allows us to update the above relation to read:

$$l_H\, \text{geo}(\mathbb{A})^{-1}\sigma_{\max}^{-m}\|x - y\|_2 \leq \left\|H\left(\phi_T^{-m}(x)\right) - H\left(\phi_T^{-m}(y)\right)\right\|_2 \leq u_H\, \text{geo}(\mathbb{A})\sigma_{\min}^{-m}\|x - y\|_2. \tag{A27}$$

From (A24), it then follows that

$$l_H^2\, \text{geo}(\mathbb{A})^{-2}\frac{\sigma_{\max}^{-2M} - 1}{\sigma_{\max}^{-2} - 1}\|x - y\|_2^2 \leq \|X_{H,T,M} - Y_{H,T,M}\|_F^2 \leq u_H^2\, \text{geo}(\mathbb{A})^2\frac{\sigma_{\min}^{-2M} - 1}{\sigma_{\min}^2 - 1}\|x - y\|_2^2. \tag{A28}$$

In turn, (A23) now implies that

$$(1 - \delta)l_H^2\, \text{geo}(\mathbb{A})^{-2}\frac{\sigma_{\max}^{-2M} - 1}{\sigma_{\max}^{-2} - 1} \leq \frac{\|F_{h_\alpha,T,M}(x) - F_{h_\alpha,T,M}(y)\|_2^2}{\|x - y\|_2^2} \tag{A29}$$

$$\leq (1 + \delta)u_H^2\, \text{geo}(\mathbb{A})^2\frac{\sigma_{\min}^{-2M} - 1}{\sigma_{\min}^{-2} - 1},$$

except with a probability of at most $C_6(\theta)\rho$. To reiterate, the above relation holds under (A20) (with $\gtrsim_\theta$ rather than $\gtrsim$), and under the mild assumption that

$$\frac{\text{vol}(\mathbb{A})^{\frac{1}{\dim(\mathbb{A})}}}{\text{rch}(\mathbb{A}_{H,T,M})} \gtrsim \frac{1}{\eta_{\min}\sqrt{\dim(\mathbb{A})}}\left(\frac{\sigma_{\max}^{-2M\dim(\mathbb{A})} - 1}{\sigma_{\max}^{-2\dim(\mathbb{A})} - 1}\right)^{-\frac{1}{2\dim(\mathbb{A})}}. \tag{A30}$$

As our last step, we now remove the stable rank from the right hand side of (A20). To accomplish that, we focus on the requirement that

$$R_{H,T,M}(\mathbb{A}) \gtrsim_\theta \delta^{-2}\dim(\mathbb{A})\log(R_{H,T,M}(\mathbb{A})). \tag{A31}$$

The *Lambert W-function* $W(\cdot)$ [65, §4.13] is defined through the relation $W(z)e^{W(z)} = z$. Strictly speaking, the Lambert W-function is not a function, as it is multi-valued when $z < 0$. In this case, $W(z)$ denotes the preimage of $W(z)e^{W(z)} = z$. Then, it is not difficult to verify that the requirement above is equivalent to

$$R_{H,T,M}(\mathbb{A}) \gtrsim_\theta e^{-\min W\left(-\frac{\delta^2}{\dim(\mathbb{A})}\right)}. \tag{A32}$$

43

This allows us to rewrite (A20) as

$$
\mathrm{R}_{H,T,M}\left(\mathbb{A}\right)
$$

$$
\gtrsim_\theta \max\left[\delta^{-2}\dim(\mathbb{A})\log\left(\eta_{\max}\sqrt{\dim(\mathbb{A})}\left(\frac{\sigma_{\min}^{-2M\dim(\mathbb{A})}-1}{\sigma_{\min}^{-2\dim(\mathbb{A})}-1}\right)^{\frac{1}{2\dim(\mathbb{A})}}\frac{\mathrm{vol}\left(\mathbb{A}\right)^{\frac{1}{\dim(\mathbb{A})}}}{\mathrm{rch}(\mathbb{A}_{H,T,M})}\right)\right.
$$

$$
\left.,e^{-\min W\left(\frac{-\delta^2}{\dim(\mathbb{A})}\right)},\delta^{-2}\log\left(\frac{1}{\rho}\right)\right]. \tag{A33}
$$

The proof of Theorem III.1 is now complete.

Remark A.1. **(Poor geodesic regularity)** If the geodesic regularity of the attractor is poor (i.e., if geo($\mathbb{A}$) in (5) is large), also, if the singular values have a high ratio (as in a chaotic system), then perhaps the following slightly weaker result is more useful. Theorem III.1 holds verbatim but with the following replacing (20):

$$
1-\delta\leq\frac{\|F_{h_\alpha,T,M}(x)-F_{h_\alpha,T,M}(y)\|_F^2}{\sum_{m=0}^{M-1}\left\|H\left(\phi_T^{-m}(x)\right)-H\left(\phi_T^{-m}(y)\right)\right\|_2^2}\leq 1+\delta, \qquad \forall x,y\in\mathbb{A}. \tag{A34}
$$

Appendix B: Proof of Lemma A.1

Recall that $\mathbb{A}\subset\mathbb{R}^N$ is a bounded, boundary-less, and smooth manifold. Also, both $\phi_T:\mathbb{A}\to\mathbb{A}$ and $H:\mathbb{A}\to H(\mathbb{A})$ are diffeomorphisms. It follows that $\mathbb{A}_{H,T,M}\subset\mathbb{R}^{MP}$ (defined in (A10)) too is a bounded, boundary-less, and smooth manifold, and that $\dim(\mathbb{A}_{H,T,M})=\dim(\mathbb{A})$.

As for vol($\mathbb{A}_{H,T,M}$), we argue as follows. For $x\in\mathbb{A}$, let $DH(x):\mathbb{T}_x\mathbb{A}\to\mathbb{T}_{H(x)}H(\mathbb{A})$ be the derivative of $H(\cdot)$ at $x\in\mathbb{A}$ (see Section II C). Each tangent space may be identified with $\mathbb{R}^{\dim(\mathbb{A})}$ and, consequently, $DH(x)$ may be identified with a $\dim(\mathbb{A})\times\dim(\mathbb{A})$ matrix. By assumption, the singular spectrum of $DH(x)$ belongs to the interval $[\eta_{\min},\eta_{\max}]\subset(0,\infty)$ (see Section III A). Then, the volume element of $\mathbb{A}$ under $H(\cdot)$ deforms as

$$
\eta_{\min}^{\dim(\mathbb{A})}d\,\mathrm{vol}(x)\leq d\,\mathrm{vol}\left(H(x)\right)\leq\eta_{\max}^{\dim(\mathbb{A})}d\,\mathrm{vol}(x), \qquad \forall x\in\mathbb{A}. \tag{B1}
$$

Similarly, let $D\phi_T(x):\mathbb{T}_x\mathbb{A}\to\mathbb{T}_{\phi(x)}\mathbb{A}$ be the derivative of the flow at $x\in\mathbb{A}$. By assumption, the singular spectrum of $D\phi(x)$ belongs to the interval $[\sigma_{\min},\sigma_{\max}]\subset(0,\infty)$. Then, the volume element of $\mathbb{A}$ under $\phi_T^{-1}(\cdot)$ deforms as

$$
\sigma_{\max}^{-\dim(\mathbb{A})}d\,\mathrm{vol}(x)\leq d\,\mathrm{vol}\left(\phi_T^{-1}(x)\right)\leq\sigma_{\min}^{-\dim(\mathbb{A})}d\,\mathrm{vol}(x), \qquad \forall x\in\mathbb{A}. \tag{B2}
$$

Predicated on the above observations, we have that

$$\mathrm{vol}\left(\mathbb{A}_{H,T,M}\right) = \int_{x\in\mathbb{A}} d\,\mathrm{vol}\left(x_{H,T,M}\right) \qquad \text{(see (A10))}$$

$$= \int_{x\in\mathbb{A}} \sqrt{\sum_{m=0}^{M-1} d\,\mathrm{vol}\left(H\left(\phi_T^{-m}\left(x\right)\right)\right)^2} \qquad \text{(see (A9))}$$

$$\leq \int_{x\in\mathbb{A}} \sqrt{\eta_{\max}^{2\dim(\mathbb{A})} \sum_{m=0}^{M-1} \sigma_{\min}^{-2m\dim(\mathbb{A})} d\,\mathrm{vol}\left(x\right)} \qquad \text{(see (B1),(B2))}$$

$$= \eta_{\max}^{\dim(\mathbb{A})} \int_{x\in\mathbb{A}} \sqrt{\frac{\sigma_{\min}^{-2M\dim(\mathbb{A})} - 1}{\sigma_{\min}^{-2\dim(\mathbb{A})} - 1}} d\,\mathrm{vol}\left(x\right)$$

$$= \eta_{\max}^{\dim(\mathbb{A})} \sqrt{\frac{\sigma_{\min}^{-2M\dim(\mathbb{A})} - 1}{\sigma_{\min}^{-2\dim(\mathbb{A})} - 1}}\, \mathrm{vol}\left(\mathbb{A}\right).$$

A similar argument establishes that

$$\mathrm{vol}\left(\mathbb{A}_{H,T,M}\right) \geq \eta_{\min}^{\dim(\mathbb{A})} \sqrt{\frac{\sigma_{\max}^{-2M\dim(\mathbb{A})} - 1}{\sigma_{\max}^{-2\dim(\mathbb{A})} - 1}}\, \mathrm{vol}\left(\mathbb{A}\right).$$

Appendix C: Proof of Theorem III.2 (Stable Takens' Theorem for Strange Attractors)

The proof follows the same arguments as outlined in Appendix A. We define the set $\mathfrak{Z}$ as in (A4), and we aim to control the supremum in (A3) by invoking Proposition A.1. To invoke this proposition, we must compute $d_F(\mathfrak{Z})$, $d_2(\mathfrak{Z})$, and $\gamma_2\left(\mathfrak{Z}, \|\cdot\|\right)$. As in (A6), we have

$$d_F\left(\mathfrak{Z}\right) = 1,$$

and as in (A7), we have

$$d_2\left(\mathfrak{Z}\right) = \frac{1}{\sqrt{\mathrm{R}_{H,T,M}\left(\mathbb{A}\right)}}.$$

45

To bound $\gamma_2\left(\mathfrak{Z},\|\cdot\|\right)$, we have

$$
\begin{aligned}
\gamma_2&\left(\mathfrak{Z},\|\cdot\|\right)\\
&\lesssim \int_0^\infty \sqrt{\log\left(\#\left(\mathfrak{Z},\|\cdot\|,s\right)\right)}\,ds && \text{(see (A5))}\\
&= \int_0^{2d_2(\mathfrak{Z})} \sqrt{\log\left(\#\left(\mathfrak{Z},\|\cdot\|,s\right)\right)}\,ds && \text{(from (A7): } s\geq 2d_2\left(\mathfrak{Z}\right)\implies \#\left(\mathfrak{Z},\|\cdot\|,s\right)=1\text{)}\\
&\leq \int_0^{\frac{2}{\sqrt{\mathrm{R}_{H,T,M}(\mathbb{A})}}} \sqrt{\log\left(\#\left(\mathfrak{Z},\|\cdot\|,s\right)\right)}\,ds && \text{(see (A7))}\\
&\leq \int_0^{\frac{2}{\sqrt{\mathrm{R}_{H,T,M}(\mathbb{A})}}} \sqrt{\log\left(\#\left(U\left(\mathbb{A}_{H,T,M}\right),\|\cdot\|_2,s\right)\right)}\,ds && \text{(see (A12))} && \text{(C1)}
\end{aligned}
$$

This allows us to focus on estimating the covering number of $U\left(\mathbb{A}_{H,T,M}\right)$. The following lemma is proved in Appendix D.

Lemma C.1. *Under the assumptions of Theorem III.2, for all $0 < s < 2$,*

$$
\#\left(U\left(\mathbb{A}_{H,T,M}\right),\|\cdot\|_2,s\right)\leq \left(\frac{\Delta}{s}\right)^{7\,\mathrm{tandim}(\mathbb{A}_{H,T,M})}, \tag{C2}
$$

where

$$
\Delta = \sqrt{24\sqrt{MP}\,\mathrm{tancov}(\mathbb{A}_{H,T,M})\left(\mathrm{cov}(\mathbb{A}_{H,T,M})\right)^{1/\,\mathrm{boxdim}(\mathbb{A}_{H,T,M})}}.
$$

Now, with (C2), we may further bound the right hand side of (C1). Omitting some intermediate steps, we conclude that

$$
\gamma_2\left(\mathfrak{Z},\|\cdot\|\right)\lesssim \sqrt{\frac{7\,\mathrm{tandim}(\mathbb{A}_{H,T,M})}{\mathrm{R}_{H,T,M}(\mathbb{A})}\log\left(\Delta^2\,\mathrm{R}_{H,T,M}(\mathbb{A})\right)} \tag{C3}
$$

as long as $\mathrm{R}_{H,T,M}(\mathbb{A})\geq 16\Delta^{-2}$. (This is guaranteed since $\Delta\geq 4$.) Then, if we assume that

$$
\mathrm{R}_{H,T,M}(\mathbb{A})\gtrsim_\theta \delta^{-2}7\,\mathrm{tandim}(\mathbb{A}_{H,T,M})\log\left(\Delta^2\,\mathrm{R}_{H,T,M}(\mathbb{A})\right)+\delta^{-2}\log\left(\frac{1}{\rho}\right), \tag{C4}
$$

we can guarantee that the following inequalities hold:

$$
d_2(\mathfrak{Z})=\frac{1}{\sqrt{\mathrm{R}_{H,T,M}(\mathbb{A})}}\lesssim_\theta \delta\left(\log\left(\frac{1}{\rho}\right)\right)^{-\frac{1}{2}} \tag{C5}
$$

and

$$
\gamma_2\left(\mathfrak{Z},\|\cdot\|\right)\lesssim_\theta \delta. \tag{C6}
$$

Subsequently, the quantities E_1, E_2, and E_3 may be bounded as in (A21) (with $\lesssim$ replaced by $\lesssim_\theta$ throughout), and finally applying Proposition A.1 with $v = \delta$ yields

$$\mathbb{P}\left[\sup_{x,y\in\mathbb{A}}\left|\frac{\|F_{h_\alpha,T,M}(x) - F_{h_\alpha,T,M}(y)\|_2^2}{\|X_{H,T,M} - Y_{H,T,M}\|_F^2} - 1\right| > \delta\right] \le C_7(\theta)\rho, \tag{C7}$$

for $C_7(\theta)$ that depends only on θ. This gives one conclusion (analogous to Remark A.1), which may be of some value: with probability at least $1 - \rho$, (24) holds for all $x, y \in \mathbb{A}$ with $x \ne y$.

We may further strengthen this conclusion by following the remaining steps in Appendix A. If we suppose that (25) holds, then we can use the bi-Lipschitz property of $H(\cdot)$ in Section III A to conclude (A27), (A28), and thus (26).

Finally, as in Appendix A, we can remove the stable rank from the right hand side of (C4) using the Lambert W-function to obtain (23).

Appendix D: Proof of Lemma C.1

To bound the covering number of $U(\mathbb{A}_{H,T,M})$, we start by defining the sets of long and short chords as

$$U_\gamma^l = \left\{\frac{a - b}{\|a - b\|_2} \; : \; a, b \in U(\mathbb{A}_{H,T,M}), \; \|a - b\|_2 > \gamma\right\},$$

$$U_\gamma^s = \left\{\frac{a - b}{\|a - b\|_2} \; : \; a, b \in U(\mathbb{A}_{H,T,M}), \; \|a - b\|_2 \le \gamma\right\},$$

where $\gamma > 0$ is a parameter to be set below. Noting that $U(\mathbb{A}_{H,T,M}) = U_\gamma^l \cup U_\gamma^s$, it suffices to bound the covering numbers of U_γ^l and U_γ^s separately.

We first bound the covering number of U_γ^l. Let K' denote a $\left(\frac{\gamma s}{8}, \|\cdot\|_2\right)$-cover of $\mathbb{A}_{H,T,M}$. To each point in K' (which has distance $\frac{\gamma s}{4}$ or less from $\mathbb{A}_{H,T,M}$), we may associate its closest point that belongs to $\mathbb{A}_{H,T,M}$. Gathering these points, we obtain a new covering we will denote by K such that $\#K \le \#K'$, $K \subset \mathbb{A}_{H,T,M}$, and K is a $\left(\frac{\gamma s}{4}, \|\cdot\|_2\right)$-cover of $\mathbb{A}_{H,T,M}$.

Now, for an arbitrary $\frac{a-b}{\|a-b\|_2} \in U_\gamma^l$, we have $\|a - b\|_2 > \gamma$ by the definition of U_γ^l. Also, by the covering construction above, there exist points $a', b' \in K$ such that

$$\|a - a'\|_2 \le \frac{\gamma s}{4} \quad \text{and} \quad \|b - b'\|_2 \le \frac{\gamma s}{4}.$$

Now, consider the Euclidean distance between $\frac{a-b}{\|a-b\|_2}$ and $\frac{a'-b'}{\|a'-b'\|_2}$. Following the proof techniques of Lemma 4.1 in [34], we have

$$
\begin{aligned}
\left\| \frac{a-b}{\|a-b\|_2} - \frac{a'-b'}{\|a'-b'\|_2} \right\|_2 &\leq \left\| \frac{a-b}{\|a-b\|_2} - \frac{a'-b'}{\|a-b\|_2} \right\|_2 + \left\| \frac{a'-b'}{\|a-b\|_2} - \frac{a'-b'}{\|a'-b'\|_2} \right\|_2 \\
&= \frac{\|(a-a')-(b-b')\|_2}{\|a-b\|_2} + \frac{|\|a'-b'\|_2 - \|a-b\|_2|}{\|a-b\|_2 \|a'-b'\|_2} \|a'-b'\|_2 \\
&\leq \frac{\|(a-a')-(b-b')\|_2}{\|a-b\|_2} + \frac{|\|a'-b'\|_2 - \|a-b\|_2|}{\|a-b\|_2} \\
&\leq \frac{\|(a-a')-(b-b')\|_2}{\|a-b\|_2} + \frac{\|(a-a')-(b-b')\|_2}{\|a-b\|_2} \\
&\leq 2 \frac{(\|a-a'\|_2 + \|b-b'\|_2)}{\|a-b\|_2} \\
&< 2\gamma^{-1} \frac{\gamma s}{2} = s,
\end{aligned}
$$

where the triangle and inverse triangle inequality were used several times. Since the choice of $\frac{a-b}{\|a-b\|_2} \in U_\gamma^l$ was arbitrary, it follows that the set

$$
\left\{ \frac{a'-b'}{\|a'-b'\|_2} \; : \; a', b' \in K \right\}
$$

is an $(s, \|\cdot\|_2)$-cover of U_γ^l. Therefore, $\#\left(U_\gamma^l, \|\cdot\|_2, s\right) \leq (\#K)^2 \leq (\#K')^2$, where it remains to bound $\#K'$. Recalling the definitions of box-counting dimension and covering regularity, we know that $\mathbb{A}_{H,T,M}$ can be covered with cubes such that

$$
\mathcal{N}(\mathbb{A}_{H,T,M}, \zeta) \leq \text{cov}(\mathbb{A}_{H,T,M}) \zeta^{-\text{boxdim}(\mathbb{A}_{H,T,M})},
$$

where we use $\mathcal{N}(\mathbb{A}_{H,T,M}, \zeta)$ to denote the number of boxes or cubes of size ζ that intersect $\mathbb{A}_{H,T,M} \subset \mathbb{R}^{MP}$. To construct a covering with Euclidean balls of radius r, one can begin with a covering of cubes with sidelength $\frac{2r}{\sqrt{MP}}$ and inscribe each of these cubes in a ball of radius r. Thus, there exists a $\left(\frac{\gamma s}{8}, \|\cdot\|_2\right)$-cover K' of $\mathbb{A}_{H,T,M}$ with

$$
\#K' \leq \mathcal{N}(\mathbb{A}_{H,T,M}, \frac{\gamma s}{4\sqrt{MP}}) \leq \text{cov}(\mathbb{A}_{H,T,M}) \left(\frac{\gamma s}{4\sqrt{MP}}\right)^{-\text{boxdim}(\mathbb{A}_{H,T,M})}. \tag{D1}
$$

Finally,

$$
\#\left(U_\gamma^l, \|\cdot\|_2, s\right) \leq (\#K')^2 \leq (\text{cov}(\mathbb{A}_{H,T,M}))^2 \left(\frac{\gamma s}{4\sqrt{MP}}\right)^{-2\,\text{boxdim}(\mathbb{A}_{H,T,M})}.
$$

We now bound the covering number of U_γ^s. The idea is to use the generalized tangent vectors of $\mathbb{A}_{H,T,M}$ to form a cover of U_γ^s. For every $d \in K$, let $\mathcal{S}_d$ denote the unit sphere in

the generalized tangent space of $\mathbb{A}_{H,T,M}$ at d, $\mathbb{T}_d\mathbb{A}_{H,T,M}$. Let $\mathcal{C}_d$ denote a $\left(\frac{s}{2}, \|\cdot\|_2\right)$-cover for $\mathcal{S}_d$, and consider the finite set

$$\mathcal{C} := \bigcup_{d \in K} \mathcal{C}_d.$$

Observe that

$$
\begin{aligned}
\#\mathcal{C} &\leq (\#K) \sup_{d \in K} \#\mathcal{C}_d \\
&\leq \mathrm{cov}(\mathbb{A}_{H,T,M}) \left(\frac{\gamma s}{4\sqrt{MP}}\right)^{-\mathrm{boxdim}(\mathbb{A}_{H,T,M})} \sup_{d \in K} \left(1 + \frac{4}{s}\right)^{\dim(\mathbb{T}_d\mathbb{A}_{H,T,M}))} \\
&\leq \mathrm{cov}(\mathbb{A}_{H,T,M}) \left(\frac{\gamma s}{4\sqrt{MP}}\right)^{-\mathrm{boxdim}(\mathbb{A}_{H,T,M})} \left(1 + \frac{4}{s}\right)^{\mathrm{tandim}(\mathbb{A}_{H,T,M})},
\end{aligned}
$$

where the second line uses $\#K \leq \#K'$, (D1), and a well-known bound on the covering number of the Euclidean ball (see, e.g., [64, Lemma 1]). The third line holds by the definition of the tangent dimension $\mathrm{tandim}(\mathbb{A}_{H,T,M})$.

Now, for an arbitrary $\frac{a-b}{\|a-b\|_2} \in U_\gamma^s$, we have $\|a - b\|_2 \leq \gamma$ by the definition of U_γ^s. Pick $d \in K$ such that $\|d - a\|_2 \leq \frac{\gamma s}{4}$. Using the triangle inequality, it follows that $\|d - b\|_2 \leq \gamma(1 + s/4)$. Thus, both a and b are within a distance of $\gamma(1 + s/4)$ from d. By the definition of tangent covering regularity, it follows that there exists $v \in \mathbb{T}_d\mathbb{A}_{H,T,M}$ such that

$$\left\| v - \frac{a - b}{\|a - b\|_2} \right\|_2 \leq \mathrm{tancov}(\mathbb{A}_{H,T,M})\gamma(1 + s/4).$$

To achieve an $(s, \|\cdot\|_2)$-cover for U_γ^s, we must keep the right-hand side of the above smaller than s. Since $s < 2$, this is guaranteed by choosing

$$\gamma = \gamma(s) = \frac{s}{3\,\mathrm{tancov}(\mathbb{A}_{H,T,M})}.$$

With this choice of γ, we have that $\mathcal{C}$ is an $(s, \|\cdot\|_2)$-cover for U_γ^s.

Adding the covering numbers for U_γ^l and U_γ^s completes the proof: for $0 < s < 2$,

$$\# \left(U\left(\mathbb{A}_{H,T,M} \right), \| \cdot \|_2, s \right)$$

$$\leq \left(\text{cov}(\mathbb{A}_{H,T,M}) \right)^2 \left(\frac{12\sqrt{MP}\,\text{tancov}(\mathbb{A}_{H,T,M})}{s^2} \right)^{2\,\text{boxdim}(\mathbb{A}_{H,T,M})}$$

$$+ \text{cov}(\mathbb{A}_{H,T,M}) \left(\frac{12\sqrt{MP}\,\text{tancov}(\mathbb{A}_{H,T,M})}{s^2} \right)^{\text{boxdim}(\mathbb{A}_{H,T,M})} \cdot \left(1 + \frac{4}{s} \right)^{\text{tandim}(\mathbb{A}_{H,T,M})}$$

$$\leq \left(\text{cov}(\mathbb{A}_{H,T,M}) \right)^2 \left(\frac{12\sqrt{MP}\,\text{tancov}(\mathbb{A}_{H,T,M})}{s^2} \right)^{2\,\text{tandim}(\mathbb{A}_{H,T,M})}$$

$$+ \text{cov}(\mathbb{A}_{H,T,M}) \left(\frac{12\sqrt{MP}\,\text{tancov}(\mathbb{A}_{H,T,M})}{s^2} \right)^{\text{tandim}(\mathbb{A}_{H,T,M})} \cdot \left(\frac{6}{s} \right)^{\text{tandim}(\mathbb{A}_{H,T,M})}$$

$$\leq \left(\frac{\sqrt{12\sqrt{MP}\,\text{tancov}(\mathbb{A}_{H,T,M})}\,(\text{cov}(\mathbb{A}_{H,T,M}))^{1/\,\text{boxdim}(\mathbb{A}_{H,T,M})}}{s} \right)^{6\,\text{tandim}(\mathbb{A}_{H,T,M})} \left(\frac{6}{s} \right)^{\text{tandim}(\mathbb{A}_{H,T,M})}$$

$$\leq \left(\frac{\sqrt{24\sqrt{MP}\,\text{tancov}(\mathbb{A}_{H,T,M})}\,(\text{cov}(\mathbb{A}_{H,T,M}))^{1/\,\text{boxdim}(\mathbb{A}_{H,T,M})}}{s} \right)^{7\,\text{tandim}(\mathbb{A}_{H,T,M})},$$

where the second inequality follows because $\text{tandim}(\mathbb{A}_{H,T,M}) \geq \text{boxdim}(\mathbb{A}_{H,T,M})$, because $s < 2$, and because we assume $\text{tancov}(\mathbb{A}_{H,T,M}) > \frac{3}{\sqrt{MP}}$. The third inequality follows from multiplying the two summands from the second inequality, both of which are greater than or equal to 2. The fourth inequality follows because we assume $\text{tancov}(\mathbb{A}_{H,T,M}) > \frac{3}{\sqrt{MP}}$ and $\text{cov}(\mathbb{A}_{H,T,M}) > 1$.

Appendix E: Proof of Lemma IV.1

We begin by calculating the Euclidean distances on $\mathbb{A}$. For $t_1, t_2 \in [0, 1)$, note that

$$
\begin{aligned}
&\|\gamma(t_1) - \gamma(t_2)\|_2^2 \\
&= \sum_{n=0}^{N-1} \left| e^{\mathrm{i}2\pi n t_1} - e^{\mathrm{i}2\pi n t_2} \right|_2^2 \qquad \text{(see (27))} \\
&= 4 \sum_{n=0}^{N-1} \sin^2\left(\pi n(t_1 - t_2)\right) \\
&= (2N-1)\left(1 - \frac{\mathrm{Dirichlet}_{2N-1}(t_1 - t_2)}{2N-1}\right). \qquad \text{(trigonometric identity)}
\end{aligned}
$$

$$(E1)$$

Above, for integer N', $\mathrm{Dirichlet}_{N'}(\cdot)$ is the Dirichlet kernel of width $\sim \frac{2}{N'}$, that is

$$
\mathrm{Dirichlet}_{N'}(t) := \frac{\sin\left(\pi N' t\right)}{\sin\left(\pi t\right)}, \qquad \forall t \in \mathbb{R}. \tag{E2}
$$

We recall an elementary property of the Dirichlet kernel.

Lemma E.1. [33, Lemma 13] For an integer N', let $Dirichlet_{N'}(\cdot)$ be the Dirichlet kernel as defined in (E2). Then, it holds that

$$
\frac{|Dirichlet_{N'}(t)|}{N'} \leq \begin{cases} \beta_1, & \frac{2}{N'} < |t| \leq \frac{1}{2} \\ \left(1 - \frac{(\pi N' t)^2}{40}\right) + \beta_3 t^2 & |t| \leq \frac{2}{N'}, \end{cases} \qquad \forall N' > N_m. \tag{E3}
$$

for (small) universal constants $\beta_1, \beta_3 > 0$. Here, $N_m = N_m(\beta_3)$ is a sufficiently large integer.

In light of this lemma, we may compare the geodesic and Euclidean distances between $\gamma(t_1), \gamma(t_2) \in \mathbb{A}$ by writing that

$$
\begin{aligned}
1 &\leq \frac{d_{\mathbb{A}}\left(\gamma(t_1), \gamma(t_2)\right)^2}{\|\gamma(t_1) - \gamma(t_2)\|_2^2} \\
&\leq \frac{2\pi^2}{3} N(N-1) \frac{(t_1 - t_2)^2}{1 - \frac{\mathrm{Dirichlet}_{2N-1}(t_1-t_2)}{2N-1}} \qquad \text{(see (30) and (E1))}
\end{aligned}
$$

and, consequently,

$$
\begin{aligned}
1 &\leq \frac{d_{\mathbb{A}}\left(\gamma(t_1),\gamma(t_2)\right)^2}{\|\gamma(t_1)-\gamma(t_2)\|_2^2} \\
&\leq \frac{2\pi^2}{3}N(N-1)\begin{cases} \frac{(t_1-t_2)^2}{1-\beta_1}, & |t_1-t_2|>\frac{2}{2N-1} \\[2mm] \frac{(t_1-t_2)^2}{\frac{(\pi(2N-1)(t_1-t_2))^2}{40}-\beta_3(t_1-t_2)^2}, & |t_1-t_2|\leq\frac{2}{2N-1} \end{cases} \\
&= \frac{2\pi^2}{3}N(N-1)\begin{cases} \frac{1}{1-\beta_1}, & |t_1-t_2|>\frac{2}{2N-1} \\[2mm] \frac{1}{\frac{\pi^2(2N-1)^2}{40}-\beta_3}, & |t_1-t_2|\leq\frac{2}{2N-1} \end{cases} \\
&\leq \frac{2\pi^2}{3}N(N-1)\max\left[\frac{1}{1-\beta_1},\frac{1}{\frac{\pi^2(2N-1)^2}{40}-\beta_3}\right] \\
&= \frac{2\pi^2}{3(1-\beta_1)}N(N-1), \qquad \text{(when N is large enough: $N>N_m$)} \qquad \text{(E4)}
\end{aligned}
$$

Above, N_m is a sufficiently large integer. This completes the proof of Lemma IV.1.

Appendix F: Proof of Lemma IV.2

From (16), observe that

$$
\mathrm{R}_{H,T,M}(\mathbb{A}) = \inf_{t,t'\geq 0}\frac{\|G_{t,T,M}-G_{t',T,M}\|_F^2}{\|G_{t,T,M}-G_{t',T,M}\|^2}, \qquad \text{(F1)}
$$

$$
\begin{aligned}
G_{t,T,M} &- G_{t',T,M} \\
&:= \Big[\,\gamma(t)-\gamma(t')\ \ \gamma(t-T)-\gamma(t'-T)\ \ \cdots\ \ \gamma(t-(M-1)T)-\gamma(t'-(M-1)T)\,\Big] \\
&\in \mathbb{C}^{N\times M}.
\end{aligned}
$$

where we have dropped H from the notation since $H(\cdot)$ is the identity operator throughout Section IV A. Let us first compute the Frobenius norm in (F1). Note that

$$
\|G_{t,T,M}-G_{t',T,M}\|_F^2 = \sum_{m=0}^{M-1}\|\gamma(t-mT)-\gamma(t'-mT)\|_2^2,
$$

and, consequently,

$$
\|G_{t,T,M}-G_{t',T,M}\|_F^2 = 4M\sum_{n=0}^{N-1}\sin^2\left(\pi n(t-t')\right). \qquad \text{(see (E1))} \qquad \text{(F2)}
$$

Computing the spectral norm in (F1) requires a more elaborate argument. Using (27), we may verify that

$$G_{t,T,M} - G_{t',T,M} = \text{diag}\left[\gamma(t) - \gamma(t')\right] \underbrace{\left[\gamma(0) \ \gamma(-T) \ \cdots \ \gamma(-MT)\right]}_{\widetilde{H} \in \mathbb{C}^{N \times M}}, \qquad \text{(F3)}$$

from which it immediately follows that

$$\|G_{t,T,M} - G_{t',T,M}\| \leq \|\gamma(t) - \gamma(t')\|_\infty \|\widetilde{H}\|$$

$$= \max_{n \in [0:N-1]} |\sin\left(\pi n(t - t')\right)| \, \|\widetilde{H}\|. \qquad \text{(see (27))} \qquad \text{(F4)}$$

Next, we bound the spectral norm of the *Vandermonde matrix* $\widetilde{H} \in \mathbb{C}^{N \times M}$. In particular, if $T = \frac{1}{N}$ and $M \leq N$, then $\widetilde{H}$ simply consists of the first M columns of the (unnormalized) $N \times N$ Fourier matrix. Consequently, $\|\widetilde{H}\| = \sqrt{N}$.

In general, we bound the spectral norm of $\widetilde{H}$ as follows. After some algebraic manipulation, one recognizes that the corresponding Grammian matrix $G \in \mathbb{C}^{M \times M}$ is both Hermitian and Toeplitz, and that (the magnitude of) its entries are specified as

$$|G[m, m']| := \left|\left[\widetilde{H}^* \widetilde{H}\right][m, m']\right|$$

$$= |\text{Dirichlet}_N \left((m - m')T\right)|, \qquad \forall m, m' \in [0 : M - 1]. \qquad \text{(F5)}$$

Above, $\text{Dirichlet}_N(\cdot)$ stands for the Dirichlet kernel of width $\sim \frac{2}{N}$ (see (E2)). Using the Gershgorin disc theorem, it then follows that

$$\|\widetilde{H}\|^2 \leq \sum_{m=0}^{M-1} |\text{Dirichlet}_N(mT)| = N + \sum_{m=1}^{M-1} |\text{Dirichlet}_N(mT)|, \qquad \text{(F6)}$$

Assuming that $MT \leq 1$, we may use the bound $\sin(\pi N t) \leq 1$ to further simplify (F6) as

$$\|\widetilde{H}\|^2 = N + \sum_{m=1}^{M-1} |\text{Dirichlet}_N(mT)|$$

$$\leq N + 2 \sum_{mT \leq \frac{1}{2}} \frac{1}{\sin(\pi m T)}, \qquad \text{(see (E2))}$$

and, consequently,

$$\|\widetilde{H}\|^2 \leq N + \frac{2}{\sin(\pi T)} + \frac{2}{T} \int_T^{\frac{1}{2}} \frac{1}{\sin(\pi t)} \, dt \qquad (\sin(\pi t) \text{ is increasing on } [0, 1/2])$$

$$= N + \frac{2}{\sin(\pi T)} - \frac{2}{\pi T} \log\left(\tan\left(\frac{\pi T}{2}\right)\right)$$

$$\leq N + \frac{2}{\sin(\pi T)} \log\left(e/\tan\left(\frac{\pi T}{2}\right)\right). \qquad (\sin(\pi T) \leq \pi T)$$

53

After substituting the estimate above back into (F4), we obtain that

$$
\|G_{t,T,M} - G_{t',T,M}\|^2
$$

$$
\leq \max_{n\in[0:N-1]} \sin^2\left(\pi n(t-t')\right) \|\widetilde{H}\|^2
$$

$$
\leq \max_{n\in[0:N-1]} \sin^2\left(\pi n(t-t')\right) \left(N + \frac{2}{\sin(\pi T)} \log\left(e/\tan\left(\frac{\pi T}{2}\right)\right)\right). \tag{F7}
$$

With the estimates in (F2) and (F7) in hand, we finally find that

$$
\mathrm{R}_{H,T,M}(\mathbb{A}) = \inf_{t,t'\geq 0} \frac{\|G_{t,T,M} - G_{t',T,M}\|_F^2}{\|G_{t,T,M} - G_{t',T,M}\|^2} \qquad \text{(see (F1))}
$$

$$
\geq \inf_{t,t'\geq 0} \frac{\frac{1}{N}\sum_{n=0}^{N-1}\sin^2\left(\pi n(t-t')\right)}{\max_{n\in[0:N-1]}\sin^2\left(\pi n(t-t')\right)} \frac{4MN}{N + \frac{2}{\sin(\pi T)}\log\left(e/\tan\left(\frac{\pi T}{2}\right)\right)}
$$

$$
= \inf_{|t|\leq\frac{1}{2}} \frac{\frac{1}{N}\sum_{n=0}^{N-1}\sin^2\left(\pi n t\right)}{\max_{n\in[0:N-1]}\sin^2\left(\pi n t\right)} \frac{M}{\frac{1}{4} + \frac{1}{2N\sin(\pi T)}\log\left(e/\tan\left(\frac{\pi T}{2}\right)\right)}. \tag{F8}
$$

We are now left with the task of controlling the infimum in the last line above. For a fixed $t \in [-\frac{1}{2}, \frac{1}{2}]$, observe that

$$
\frac{\frac{1}{N}\sum_{n=0}^{N-1}\sin^2\left(\pi n t\right)}{\max_{n\in[0:N-1]}\sin^2\left(\pi n t\right)}
$$

$$
= \frac{2N-1}{N} \frac{1 - \frac{\mathrm{Dirichlet}_{2N-1}(t)}{2N-1}}{\max_{n\in[0:N-1]}\sin^2\left(\pi n t\right)} \qquad \text{(trigonometric identity)}
$$

$$
\geq \frac{1 - \frac{\mathrm{Dirichlet}_{2N-1}(t)}{2N-1}}{\min\left[1, (\pi N t)^2\right]}
$$

$$
\geq \frac{1}{\min\left[1, (\pi N t)^2\right]} \begin{cases} 1 - \beta_1, & |t| > \frac{2}{2N-1} \\ \frac{(\pi(2N-1)t)^2}{40} - \beta_3 t^2, & |t| \leq \frac{2}{2N-1} \end{cases}, \qquad \text{(see Lemma E.1)}
$$

54

and, consequently,

$$\frac{\frac{1}{N}\sum_{n=0}^{N-1}\sin^2(\pi n t)}{\max_{n\in[0:N-1]}\sin^2(\pi n t)}$$

$$\geq \begin{cases} 1-\beta_1, & |t| > \frac{2}{2N-1} \\ \frac{\frac{(\pi(2N-1)t)^2}{40}-\beta_3 t^2}{(\pi N t)^2}, & |t| \leq \frac{2}{2N-1} \end{cases}$$

$$\geq \begin{cases} 1-\beta_1, & |t| > \frac{2}{2N-1} \\ \frac{1}{40}-\frac{\beta_3}{\pi^2 N^2}, & |t| \leq \frac{2}{2N-1} \end{cases}$$

$$\geq \min\left[1-\beta_1, \frac{1}{40}-\frac{\beta_3}{\pi^2 N^2}\right]$$

$$\geq \min\left[1-\beta_1, \frac{1}{80}\right]$$

$$= \frac{1}{80}, \qquad (\beta_1 \approx 0.23) \tag{F9}$$

where $\beta_1, \beta_3 > 0$ are (small) absolute constants and, in particular, $\beta_1 \approx 0.23$. The fourth and last two lines above hold for sufficiently large N: $N > N_m = N_m(\beta_3)$. The above estimate is independent of t and, by substituting in (F8), leads us to

$$\mathrm{R}_{H,T,M}(\mathbb{A}) \geq \frac{1}{80}\frac{M}{\frac{1}{4}+\frac{1}{2N\sin(\pi T)}\log\left(e/\tan\left(\frac{\pi T}{2}\right)\right)}.$$

This completes the proof of Lemma IV.2.

———————————

[1] A. Wittmann and Z. Xu, Vistas in Astronomy **31**, 127 (1988).

[2] P. Brockwell and R. Davis, *Introduction to time series and forecasting*, Lecture Notes in Statistics (Springer, 2002).

[3] A. Beuter, *Nonlinear Dynamics in Physiology and Medicine*, edited by L. Glass, M. Mackey, M. Titcombe, and A. Beuter, Interdisciplinary Applied Mathematics SE (Springer-Verlag, New York, 2003).

[4] E. Mosekilde, *Topics in Nonlinear Dynamics: Applications to Physics, Biology and Economic Systems* (World Scientific Publishing Company, 1996).

[5] W. Sulis and A. Combs, eds., *Nonlinear Dynamics in Life and Social Sciences*, NATO Science Series, Vol. 320 (IOS Press, 2001).

[6] A. Saperstein and G. Mayer-Kress, The Journal of Conflict Resolution **32**, 636 (1988).

[7] C. Grebogi and J. Yorke, eds., *Impact of Chaos on Science and Society* (United Nations University Press, 1997).

[8] M. Roe, Harvard Law Review **109**, 641 (1996).

[9] M. Wolfson, A. Puri, and M. Martelli, The Journal of Conflict Resolution **36**, 119 (1992).

[10] F. Udwadia, G. Leitmann, and L. Lambertini, Discrete Dynamics in Nature and Society **2006** (2006).

[11] H. Kantz and T. Schreiber, *Nonlinear time series analysis*, Vol. 7 (Cambridge University Press, 2004).

[12] G. Yule, Philosophical Transactions of the Royal Society of London, Series A **226**, 267 (1927).

[13] Y. Bar-Shalom, X. Li, and T. Kirubarajan, *Estimation with applications to tracking and navigation: Theory algorithms and software* (Wiley, 2004).

[14] F. Takens, in *Dynamical systems and turbulence, Warwick 1980*, Lecture Notes in Mathematics, Vol. 898, edited by D. Rand and L. Young (Springer Berlin / Heidelberg, 1981) pp. 366–381.

[15] T. Sauer, J. Yorke, and M. Casdagli, Journal of Statistical Physics **65**, 579 (1991).

[16] M. Casdagli, S. Eubank, J. Farmer, and J. Gibson, Physica D: Nonlinear Phenomena **51**, 52 (1991).

[17] A. Fraser and H. Swinney, Physical Review A **33**, 1134 (1986).

[18] A. Kazem, E. Sharifi, F. K. Hussain, M. Saberi, and O. K. Hussain, Applied Soft Computing **13**, 947 (2013).

[19] L. Uzal, G. Grinblat, and P. Verdes, Physical Review E **84**, 016223 (2011).

[20] D. Kugiumtzis, Physica D: Nonlinear Phenomena **95**, 13 (1996).

[21] W.-X. Wang, R. Yang, Y.-C. Lai, V. Kovanis, and C. Grebogi, Phys. Rev. Lett. **106** (2011).

[22] E. Balaguer-Ballester, C. C. Lapish, J. K. Seamans, and D. Durstewitz, PLoS Comput Biol **7**, e1002057 (2011).

[23] J. P. Bello, IEEE Transactions on Audio, Speech, and Language Processing **19**, 2013 (2011).

[24] F. Hamilton, T. Berry, and T. Sauer, Physical Review X **6**, 011021 (2016).

[25] S. Dasgupta and A. Gupta, Random Structures and Algorithms **22**, 60 (2002).

[26] D. Donoho, IEEE Transactions on Information Theory **52**, 1289 (2006).

[27] E. Candes and T. Tao, IEEE Transactions on Information Theory **52**, 5406 (2006).

[28] E. Candes, Comptes Rendus Mathematique **346**, 589 (2008).

[29] J. Tenenbaum, V. de Silva, and J. Langford, Science **290**, 2319 (2000).

[30] D. Broomhead and M. Kirby, Neural Computation **13**, 2595 (2001).

[31] C. Hegde, A. C. Sankaranarayanan, W. Yin, and R. Baraniuk, IEEE Transactions on Signal Processing **63**, 6109 (2015).

[32] G. Puy, M. Davies, and R. Gribonval, IEEE Transactions on Information Theory **63**, 2171 (2017).

[33] A. Eftekhari and M. Wakin, Applied and Computational Harmonic Analysis **39**, 67 (2015).

[34] K. Clarkson, in *Proceedings of 24th Annual Symposium on Computational Geometry* (ACM, 2008) pp. 39–48.

[35] H. Yap, M. Wakin, and C. Rozell, IEEE Journal of Selected Topics in Signal Processing **7**, 720 (2013).

[36] J. Stark, Journal of Nonlinear Science **9**, 255 (1999).

[37] W. Dechert and R. Gencay, Physica D: Nonlinear Phenomena **90**, 40 (1996).

[38] H. Abarbanel, R. Brown, J. Sidorowich, and L. Tsimring, Reviews of modern physics **65**, 1331 (1993).

[39] T. Asefa, M. Kemblowski, U. Lall, and G. Urroz, Water resources research **41** (2005).

[40] T. Schreiber and D. Kaplan, Chaos (Woodbury, N.Y.) **6**, 87 (1996).

[41] E. Ott and J. Yorke, Physical Review Letters **64**, 1196 (1990).

[42] L. Pecora and T. Carroll, Physical Review A **44**, 2374 (1991).

[43] C. P. Silva and A. M. Young, in *IEEE Aerospace Conference Proceedings*, Vol. 1 (2000) pp. 279–299.

[44] H. Dedieu, Circuits and Systems I, IEEE Transactions on **44**, 948 (1997).

[45] G. Sugihara, R. May, H. Ye, C.-H. Hsieh, E. Deyle, M. Fogarty, and S. Munch, Science **338**, 496 (2012).

[46] T. Schreiber and P. Grassberger, Physics letters A **160**, 411 (1991).

[47] W. Liebert, K. Pawelzik, and H. Schuster, EPL (Europhysics Letters) **14**, 521 (1991).

[48] M. Kennel, R. Brown, and H. Abarbanel, Physical review A **45**, 3403 (1992).

[49] L. Cao, Physica D: Nonlinear Phenomena **110**, 43 (1997).

[50] T. Buzug and G. Pfister, Physical review A **45**, 7073 (1992).

[51] M. Rosenstein, J. Collins, and C. De Luca, Physica D: Nonlinear Phenomena **73**, 82 (1994).

[52] L. Pecora, T. Carroll, and J. Heagy, Physical Review E **52**, 3420 (1995).

[53] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin, Constructive Approximation **28**, 253 (2008).

[54] H. Rauhut, J. Romberg, and J. Tropp, Applied and Computational Harmonic Analysis **32**, 242 (2012).

[55] F. Krahmer, S. Mendelson, and H. Rauhut, Communications on Pure and Applied Mathematics **67**, 1877 (2014).

[56] A. Eftekhari, H. L. Yap, C. J. Rozell, and M. B. Wakin, Applied and Computational Harmonic Analysis **38**, 1 (2015).

[57] T. Blumensath and M. E. Davies, IEEE Transactions on Information Theory **55**, 1872 (2009).

[58] G. Tran and R. Ward, arXiv preprint arXiv:1607.01067 (2016).

[59] H. Yap and C. Rozell, IEEE Transactions on Signal Processing **59**, 4781 (2011).

[60] M. Spivak, *A Comprehensive Introduction to Differential Geometry*, A Comprehensive Introduction to Differential Geometry No. v. 3 (Publish or Perish, Incorporated, 1975).

[61] P. Niyogi, S. Smale, and S. Weinberger, Discrete and Computational Geometry **39**, 419 (2008).

[62] H. Federer, Transactions of the American Mathematical Society , 418 (1959).

[63] J. Tropp, in *Proceedings of the Twentieth Annual ACM-SIAM Symposium on Discrete Algorithms* (Society for Industrial and Applied Mathematics, 2009) pp. 978–986.

[64] R. Vershynin, in *Compressed Sensing, Theory and Applications*, edited by Y. Eldar and G. Kutyniok (Cambridge University Press, 2012) Chap. 5, pp. 210–268.

[65] F. Olver, N. I. of Standards, and T. (U.S.), *NIST handbook of mathematical functions* (Cambridge University Press, 2010).

[66] P. Grassberger and I. Procaccia, in *The Theory of Chaotic Attractors* (Springer, 2004) pp. 170–189.

[67] W. Ott and J. Yorke, SIAM Journal on Applied Dynamical Systems , 1 (2003).

[68] M. Brin, W. Ott, and J. Yorke, Topology and its Applications **145**, 233 (2004).

[69] J. N. Kutz, *Data-driven modeling & scientific computation: Methods for complex systems & big data* (OUP Oxford, 2013).

[70] H. Jaeger and H. Haas, Science **304**, 78 (2004).

[71] L. Cao, A. Mees, and K. Judd, Physica D: Nonlinear Phenomena **121**, 75 (1998).

[72] M. Talagrand, *The generic chaining: Upper and lower bounds of stochastic processes*, Springer Monographs in Mathematics (Springer, 2006).