



On the Shroer–Sauer–Ott–Yorke Predictability Conjecture for Time-Delay Embeddings

Krzysztof Barański¹ , Yonatan Gutman², Adam Śpiewak³

¹ Institute of Mathematics, University of Warsaw, ul. Banacha 2, 02-097 Warszawa, Poland.

E-mail: baranski@mimuw.edu.pl

² Institute of Mathematics, Polish Academy of Sciences, ul. Śniadeckich 8, 00-656 Warszawa, Poland.

E-mail: y.gutman@impan.pl

³ Department of Mathematics, Bar-Ilan University, Ramat Gan 5290002, Israel.

E-mail: ad.spiewak@gmail.com

Received: 16 July 2021 / Accepted: 5 January 2022

Published online: 2 February 2022 – © The Author(s) 2022

Abstract: Shroer, Sauer, Ott and Yorke conjectured in 1998 that the Takens delay embedding theorem can be improved in a probabilistic context. More precisely, their conjecture states that if μ is a natural measure for a smooth diffeomorphism of a Riemannian manifold and k is greater than the information dimension of μ , then k time-delayed measurements of a one-dimensional observable h are generically sufficient for a *predictable* reconstruction of μ -almost every initial point of the original system. This reduces by half the number of required measurements, compared to the standard (deterministic) setup. We prove the conjecture for ergodic measures and show that it holds for a generic smooth diffeomorphism, if the information dimension is replaced by the Hausdorff one. To this aim, we prove a general version of predictable embedding theorem for injective Lipschitz maps on compact sets and arbitrary Borel probability measures. We also construct an example of a C^∞ -smooth diffeomorphism with a natural measure, for which the conjecture does not hold in its original formulation.

1. Introduction

1.1. General background. This paper concerns probabilistic aspects of the *Takens delay embedding theorem*, dealing with the problem of reconstructing a dynamical system from a sequence of measurements of a one-dimensional observable. More precisely, let $T : X \rightarrow X$ be a transformation on a *phase space* X . Fix $k \in \mathbb{N}$ and consider a function (*observable*) $h : X \rightarrow \mathbb{R}$ together with the corresponding *k-delay coordinate map*

$$\phi : X \rightarrow \mathbb{R}^k, \quad \phi(x) = (h(x), \dots, h(T^{k-1}x)).$$

Takens-type delay embedding theorems state that if k is large enough, then ϕ is an embedding (i.e. is injective) for a typical observable h . The injectivity of ϕ ensures that an (unknown) initial state $x \in X$ of the system can be uniquely recovered from the sequence of k measurements $h(x), \dots, h(T^{k-1}x)$ of the observable h , performed along

the orbit of x . It also implies that the dynamical system (X, T) has a reliable model in \mathbb{R}^k of the form $(\tilde{X}, \tilde{T}) = (\phi(X), \phi \circ T \circ \phi^{-1})$.

This line of research originates from the seminal paper of Takens [Tak81] on diffeomorphisms of compact manifolds. Extensions of Takens' work were obtained in several categories, e.g. in [SYC91, Sta99, Cab00, Rob05, Gut16, GQS18, SBDH97], [SBDH03, NV20] (see also [Rob11, BGŚ20] for a more detailed overview). A common feature of these results is that the minimal number of measurements sufficient for an exact reconstruction of the system is $k \approx 2 \dim X$, where $\dim X$ is the dimension of the phase space X . This threshold agrees with the one appearing in the classical non-dynamical embedding theorems (e.g. Whitney theorem [Whi36], Menger–Nöbeling theorem [HW41, Theorem V.2] and Mañé theorem [Rob11, Theorem 6.2]). It is worth to notice that Takens-type theorems serve as a justification of the validity of time-delay based procedures, which are actually used in applications (see e.g. [HGLS05, KY90, SGM90, SM90]) and have been met with a great interest among mathematical physicists (see e.g. [PCFS80, HBS15, SYC91, Vos03]).

In 1998, Shroer, Sauer, Ott and Yorke conjectured (see [SSOY98, Conjecture 1]), that for smooth diffeomorphisms on compact manifolds, in a probabilistic setting (i.e. when the initial point $x \in X$ is chosen randomly according to a natural probability measure μ), the number of measurements required for an almost sure *predictable* reconstruction of the system can be generically reduced by half, up to the *information dimension* of μ . A precise formulation is given below in Sect. 1.2. We will refer to this conjecture as *Shroer–Sauer–Ott–Yorke predictability conjecture* or *SSOY predictability conjecture*. In [SSOY98], the authors provided some heuristic arguments supporting the conjecture together with its numerical verification for some examples (Hénon and Ikeda maps). However, a rigorous proof of the conjecture has been unknown up to now.

In this paper, we prove a general version of a predictable embedding theorem (Theorem 1.7), valid for injective Lipschitz transformations of compact sets and arbitrary Borel probability measures, which shows that an almost sure predictable reconstruction of the system is possible with the number of measurements reduced to the *Hausdorff dimension* of μ , under a mild assumption bounding the dimensions of sets of periodic points of low periods. As a corollary, we obtain the SSOY predictability conjecture for generic smooth C^r -diffeomorphisms on compact manifolds for $r \geq 1$, with information dimension replaced by the Hausdorff one (Corollary 1.9) and the original conjecture for arbitrary C^r -diffeomorphisms and ergodic measures (Corollary 1.10). We also construct an example of a C^∞ -smooth diffeomorphism of a compact Riemannian manifold with a non-ergodic natural measure, for which the original conjecture does not hold (Theorem 1.11). This shows that in a general case, the change of the information dimension to the Hausdorff one is necessary.

Let us note that the SSOY predictability conjecture has been invoked in a number of papers (see e.g. [Liu10, MS04, OL98]) as a theoretical argument for reducing the number of measurements required for a reliable reconstruction of the system, also in applications (see e.g. [QMAV99] studying neural brain activity in focal epilepsy). Our result provides a mathematically rigorous proof of the correctness of these procedures.

1.2. Shroer–Sauer–Ott–Yorke predictability conjecture. Before we formulate the conjecture stated in [SSOY98] in a precise way, we need to introduce some preliminaries, in particular the notion of predictability. In the sequel, we consider a general situation, when the phase space X is an arbitrary compact set in \mathbb{R}^N (note that by the Whitney

embedding theorem [Whi36], we can assume that a smooth compact manifold is embedded in \mathbb{R}^N for sufficiently large N). We denote the (topological) support of a measure μ by $\text{supp } \mu$ and write $\phi_*\mu$ for a push-forward of μ by a measurable transformation ϕ , defined by $\phi_*\mu(A) = \mu(\phi^{-1}(A))$ for measurable sets A .

Definition 1.1. Let $X \subset \mathbb{R}^N$ be a compact set, let μ be a Borel probability measure with support in X and let $T: X \rightarrow X$ be a Borel transformation (i.e. such that the preimage of any Borel set is Borel). Fix $k \in \mathbb{N}$. Let $h: X \rightarrow \mathbb{R}$ be a Borel observable and let $\phi: X \rightarrow \mathbb{R}^k$ given by $\phi(x) = (h(x), \dots, h(T^{k-1}x))$ be the corresponding k -delay coordinate map. Set $\nu = \phi_*\mu$ (considered as a Borel measure in \mathbb{R}^k) and note that $\text{supp } \nu \subset \phi(X)$. For $y \in \text{supp } \nu$ and $\varepsilon > 0$ define

$$\chi_\varepsilon(y) = \frac{1}{\mu(\phi^{-1}(B(y, \varepsilon)))} \int_{\phi^{-1}(B(y, \varepsilon))} \phi(Tx) d\mu(x),$$

$$\sigma_\varepsilon(y) = \left(\frac{1}{\mu(\phi^{-1}(B(y, \varepsilon)))} \int_{\phi^{-1}(B(y, \varepsilon))} \|\phi(Tx) - \chi_\varepsilon(y)\|^2 d\mu(x) \right)^{\frac{1}{2}},$$

where $B(y, \varepsilon)$ denotes the open ball of radius ε centered at y . In other words, $\chi_\varepsilon(y)$ is the conditional expectation of the random variable $\phi \circ T$ (with respect to μ) given $\phi \in B(y, \varepsilon)$, while $\sigma_\varepsilon(y)$ is its conditional standard deviation. Define also the *prediction error* at y as

$$\sigma(y) = \lim_{\varepsilon \rightarrow 0} \sigma_\varepsilon(y),$$

provided the limit exists. A point y is said to be *predictable* if $\sigma(y) = 0$.

Note that the prediction error depends on the observable h . We simplify the notation by suppressing this dependence.

Remark 1.2. Note that the predictability of points of the support of the measure ν does not imply that the delay coordinate map ϕ is injective. Indeed, if h (and hence ϕ) is constant, then every point $y \in \text{supp } \nu$ is predictable.

Remark 1.3. (Farmer and Sidorowich algorithm) As explained in [SSOY98], the notion of predictability arises naturally in the context of a prediction algorithm proposed by Farmer and Sidorowich in [FS87]. To describe it, suppose that for a point $x \in X$ we are given a sequence of measurements $h(x), \dots, h(T^{n+k-1}x)$ of the observable h for some $n \in \mathbb{N}$. This defines a sequence of k -delay coordinate vectors of the form

$$y_i = (h(T^i x), \dots, h(T^{i+k-1} x)), \quad i = 0, \dots, n.$$

Knowing the sample values of y_0, \dots, y_n , we would like to predict the one-step future of the model, i.e. the value of the next point $y_{n+1} = (h(T^{n+1}x), \dots, h(T^{n+k}x))$. For a small $\varepsilon > 0$ we define the predicted value of y_{n+1} as

$$\widehat{y}_{n+1} = \frac{1}{\#\mathcal{I}} \sum_{i \in \mathcal{I}} y_{i+1} \quad \text{for } \mathcal{I} = \{0 \leq i < n : y_i \in B(y_n, \varepsilon)\}.$$

In other words, the predicted value of y_{n+1} is taken to be the average of the values y_{i+1} , where we count only those i , for which y_i are ε -close to the last known point y_n .

Notice that if the k -delay coordinate map ϕ is an embedding, then the points y_i form an orbit of y_0 under the model transformation \tilde{T} defined by the delay coordinate map ϕ , i.e. $y_i = \tilde{T}^i(y_0)$ for $(\tilde{X}, \tilde{T}) = (\phi(X), \phi \circ T \circ \phi^{-1})$. Hence, in this case the predicted value $y_{n+1} = \tilde{T}(y_n)$ is the average of the values $y_{i+1} = \tilde{T}(y_i)$, $i \in \mathcal{I}$.

If the initial point $x \in X$ is chosen randomly according to an ergodic probability measure μ , then for $n \rightarrow \infty$, the collection of points y_i , $i \in \mathcal{I}$ is asymptotically distributed in $B(y_n, \varepsilon)$ according to the measure $\nu = \phi_*\mu$. Therefore, the value of $\sigma_\varepsilon(y_n)$ from Definition 1.1 approaches asymptotically the standard deviation of the predicted point $\widehat{y_{n+1}}$. The condition of predictability states that this standard deviation converges to zero as ε tends to zero.

In [SSOY98], the Shroer–Sauer–Ott–Yorke predictability conjecture is stated for a special class of measures, called *natural measures*. To define it, recall first that a measure μ on X is *invariant* for a measurable map $T : X \rightarrow X$ if $\mu(T^{-1}(A)) = \mu(A)$ for every measurable set $A \subset X$. A set $\Lambda \subset X$ is called *T-invariant* if $T(\Lambda) \subset \Lambda$.

Definition 1.4. Let X be a compact Riemannian manifold and $T : X \rightarrow X$ be a smooth diffeomorphism. A compact T -invariant set $\Lambda \subset X$ is called an *attractor*, if the set $B(\Lambda) = \{x \in X : \lim_{n \rightarrow \infty} \text{dist}(T^n x, \Lambda) = 0\}$ is an open set containing Λ . The set $B(\Lambda)$ is called the *basin of attraction* to Λ . A T -invariant Borel probability measure μ on Λ is called a *natural measure* if

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} \delta_{T^i x} = \mu$$

for almost every $x \in B(\Lambda)$ with respect to the volume measure on X , where δ_y denotes the Dirac measure at y and the limit is taken in the weak-* topology.

Remark 1.5. Note that in ergodic theory of dynamical systems, some authors use the name *physical measure* or *SRB (Sinai–Ruelle–Bowen) measure* for similar concepts (see e.g. [You02]). The term ‘natural measure’ occurs commonly in mathematical physics literature (see e.g. [Ott02, OY08]).

Definition 1.6. For a Borel probability measure μ in \mathbb{R}^N with compact support define its *lower* and *upper information dimensions* as ¹

$$\begin{aligned} \underline{\text{ID}}(\mu) &= \liminf_{\varepsilon \rightarrow 0} \int_{\text{supp } \mu} \frac{\log \mu(B(x, \varepsilon))}{\log \varepsilon} d\mu(x), \\ \overline{\text{ID}}(\mu) &= \limsup_{\varepsilon \rightarrow 0} \int_{\text{supp } \mu} \frac{\log \mu(B(x, \varepsilon))}{\log \varepsilon} d\mu(x). \end{aligned}$$

If $\underline{\text{ID}}(\mu) = \overline{\text{ID}}(\mu)$, then we denote their common value as $\text{ID}(\mu)$ and call it the *information dimension* of μ .

¹ Information dimensions are often defined in an equivalent way as

$$\underline{\text{ID}}(\mu) = \liminf_{\varepsilon \rightarrow 0} \frac{1}{\log \varepsilon} \sum_{C \in \mathcal{C}_\varepsilon} \mu(C) \log \mu(C), \quad \overline{\text{ID}}(\mu) = \limsup_{\varepsilon \rightarrow 0} \frac{1}{\log \varepsilon} \sum_{C \in \mathcal{C}_\varepsilon} \mu(C) \log \mu(C),$$

where \mathcal{C}_ε is the partition of \mathbb{R}^N into cubes with side lengths ε and vertices in the lattice $(\varepsilon\mathbb{Z})^N$ (see e.g. [WV10, Appendix I]).

We are now ready to state the SSOY predictability conjecture in its original form as stated in [SSOY98]. Recall that for a map $T : X \rightarrow X$ with a Borel probability measure μ , a number $k \in \mathbb{N}$ and a function $h : X \rightarrow \mathbb{R}$, we consider the k -delay coordinate map for the observable h defined by

$$\phi(x) = \phi_{h,k}(x) = (h(x), \dots, h(T^{k-1}x)).$$

To emphasize the dependence on h and k , we will write $\phi_{h,k}$ for ϕ and $\nu_{h,k}$ for the push-forward measure $\nu = \nu_{h,k} = (\phi_{h,k})_*\mu$.

SSOY predictability conjecture ([SSOY98, Conjecture 1]) *Let $T : X \rightarrow X$ be a smooth diffeomorphism of a compact Riemannian manifold X and let $\Lambda \subset X$ be an attractor of T with a natural measure μ such that $\text{ID}(\mu) = D$. Fix $k > D$. Then $\nu_{h,k}$ -almost every point of \mathbb{R}^k is predictable for a generic observable $h : X \rightarrow \mathbb{R}$.*

Note that in this formulation some details (e.g. the type of genericity and the smoothness class of the dynamics) are not specified precisely.

1.3. Main results. Now we present the main results of the paper. First, we state a predictable embedding theorem, which holds in a general context of injective Lipschitz maps T on a compact set $X \subset \mathbb{R}^N$ equipped with a Borel probability measure μ . Recall that by the Whitney embedding theorem [Whi36], we can assume that a smooth compact manifold is embedded in \mathbb{R}^N for sufficiently large N . Our observation is that in this generality, the predictability holds if we replace the information dimension $\text{ID}(\mu)$ by the Hausdorff dimension $\dim_H \mu$ (see Sect. 2.1 for definition).

In the presented results, we understand the genericity of the observable h in the sense of *prevalence* in the space $\text{Lip}(X)$ of Lipschitz observables $h : X \rightarrow \mathbb{R}$ (with a polynomial *probe set*), which is an analogue of the ‘Lebesgue almost sure’ condition in infinite dimensional spaces (see Sect. 2.2 for precise definitions). In particular, the genericity of h holds also in the sense of prevalence in the space of C^r -smooth observables $h : X \rightarrow \mathbb{R}$, for $r \geq 1$. Let us note that it is standard to use prevalence as a notion of genericity in the context of Takens-type embedding theorems (see e.g. [SYC91, Rob11]).

It is known that Takens-type theorems require some bounds on the size of sets of T -periodic points of low periods. Following [BGŚ20], we assume $\dim_H(\mu|_{\text{Per}_p(T)}) < p$ for $p = 1, \dots, k - 1$, where

$$\text{Per}_p(T) = \{x \in X : T^p x = x\}.$$

With these remarks, our main result is the following.

Theorem 1.7 (Predictable embedding theorem for Lipschitz maps). *Let $X \subset \mathbb{R}^N$ be a compact set, let μ be a Borel probability measure on X and let $T : X \rightarrow X$ be an injective Lipschitz map. Take $k > \dim_H \mu$ and assume $\dim_H(\mu|_{\text{Per}_p(T)}) < p$ for $p = 1, \dots, k - 1$. Then for a prevalent set of Lipschitz observables $h : X \rightarrow \mathbb{R}$, the k -delay coordinate map $\phi_{h,k}$ is injective on a Borel set of full μ -measure, and $\nu_{h,k}$ -almost every point of \mathbb{R}^k is predictable.*

Remark 1.8. Notice that except of predictability, we obtain almost sure injectivity of the delay coordinate map, which means that the system can be reconstructed in \mathbb{R}^k in a one-to-one fashion on a set of full measure.

An extended version of Theorem 1.7 is proved in Sect. 3 as Theorem 3.1.

Note that the assumption on the dimension of μ restricted to the set of p -periodic points can be omitted if there are only finitely many periodic points of given period. By the Kupka–Smale theorem (see [PdM82, Chapter 3, Theorem 3.6]), the latter condition is generic (in the Baire category sense) in the space of C^r -diffeomorphisms, $r \geq 1$, of a compact manifold, equipped with the uniform C^r -topology (see [BGŚ20] for more details). Therefore, we immediately obtain the SSOY predictability conjecture for generic smooth C^r -diffeomorphisms, with information dimension replaced by the Hausdorff one.

Corollary 1.9 (SSOY predictability conjecture for generic diffeomorphisms). *Let X be a compact Riemannian manifold and $r \geq 1$. Then for a C^r -generic diffeomorphism $T: X \rightarrow X$ with a natural measure μ (or, more generally, any Borel probability measure) and $k > \dim_H \mu$, for a prevalent set (depending on T) of Lipschitz observables $h: X \rightarrow \mathbb{R}$, the k -delay coordinate map $\phi_{h,k}$ is injective on a set of full μ -measure, and $v_{h,k}$ -almost every point of \mathbb{R}^k is predictable.*

Suppose now the measure μ in Theorem 1.7 is T -invariant and ergodic. Then we have $\dim_H \mu \leq \underline{\text{ID}}(\mu) \leq \overline{\text{ID}}(\mu)$ (see Proposition 2.1). Moreover, either the set of T -periodic points has μ -measure zero, or μ is supported on a periodic orbit of T (see the proof of [BGŚ20, Remark 4.4(c)]). Hence, the assumption on the dimension of μ restricted to the set of p -periodic points can again be omitted. This proves the original SSOY conjecture for arbitrary C^r -diffeomorphisms and ergodic measures.

Corollary 1.10 (SSOY predictability conjecture for ergodic measures). *Let X be a compact Riemannian manifold, $r \geq 1$, and let $T: X \rightarrow X$ be a C^r -diffeomorphism with an ergodic natural measure μ (or, more generally, any T -invariant ergodic Borel probability measure). Take $k > \underline{\text{ID}}(\mu)$. Then for a prevalent set of Lipschitz observables $h: X \rightarrow \mathbb{R}$, the k -delay coordinate map $\phi_{h,k}$ is injective on a set of full μ -measure, and $v_{h,k}$ -almost every point of \mathbb{R}^k is predictable.*

Our final result is that the SSOY predictability conjecture does not hold in its original formulation for all smooth diffeomorphisms, i.e. the condition $k > \text{ID}(\mu)$ is not sufficient for almost sure predictability for generic observables, even if μ is within the class of natural measures.

Theorem 1.11. *There exists a C^∞ -smooth diffeomorphism of the 3-dimensional compact Riemannian manifold $X = \mathbb{S}^2 \times \mathbb{S}^1$ with a natural measure μ , such that $\text{ID}(\mu) < 1$ and for a prevalent set of Lipschitz observables $h: X \rightarrow \mathbb{R}$, there exists a positive $v_{h,1}$ -measure set of non-predictable points. In particular, the set of Lipschitz observables $h: X \rightarrow \mathbb{R}$ for which $v_{h,1}$ -almost every point of \mathbb{R}^k is predictable, is not prevalent.*

The construction is presented in Sect. 4 (see Theorem 4.14 for details).

Remark 1.12. Theorem 1.11 shows that the original SSOY predictability conjecture fails for a specific system (X, T) . It remains an open question whether it holds for a generic C^r -diffeomorphism T of a given compact Riemannian manifold X . By Corollary 1.9, this would follow from the dimension conjecture of Farmer, Ott and Yorke [FOY83, Conjecture 1], which (in particular) states that the Hausdorff and information dimension of the natural measure typically coincide.

Organization of the paper. Section 2 contains preliminary material, gathering definitions and tools required for the rest of the paper. Theorem 1.7 and its extension Theorem 3.1 are proved in Sect. 3. Section 4 contains a construction of the example presented in Theorem 1.11, divided into several steps.

2. Preliminaries

2.1. Hausdorff and information dimensions. For $s > 0$, the s -dimensional (outer) Hausdorff measure of a set $X \subset \mathbb{R}^N$ is defined as

$$\mathcal{H}^s(X) = \lim_{\delta \rightarrow 0} \inf \left\{ \sum_{i=1}^{\infty} |U_i|^s : X \subset \bigcup_{i=1}^{\infty} U_i, |U_i| \leq \delta \right\},$$

where $|\cdot|$ denotes the diameter of a set (with respect to the Euclidean distance in \mathbb{R}^N). The Hausdorff dimension of X is given as

$$\dim_H X = \inf\{s > 0 : \mathcal{H}^s(X) = 0\} = \sup\{s > 0 : \mathcal{H}^s(X) = \infty\}.$$

The (upper) Hausdorff dimension of a finite Borel measure μ in \mathbb{R}^N is defined as

$$\dim_H \mu = \inf\{\dim_H X : X \subset \mathbb{R}^N \text{ is a Borel set of full } \mu\text{-measure}\}.$$

By the Whitney embedding theorem [Whi36], we can assume that a smooth compact manifold is smoothly embedded in the Euclidean space, hence the Hausdorff dimension is well defined also for Borel measures on manifolds.

In general, $\underline{\text{ID}}(\mu)$ and $\overline{\text{ID}}(\mu)$ are not comparable with $\dim_H \mu$ (see [FLR02, Sect. 3]). One can however obtain inequalities between them for measures which are ergodic with respect to Lipschitz transformations.

Proposition 2.1. *Let $X \subset \mathbb{R}^N$ be a closed set, let $T : X \rightarrow X$ be a Lipschitz map and let μ be a T -invariant and ergodic Borel probability measure on X . Then*

$$\dim_H \mu \leq \underline{\text{ID}}(\mu) \leq \overline{\text{ID}}(\mu).$$

Proof. The inequality $\underline{\text{ID}}(\mu) \leq \overline{\text{ID}}(\mu)$ is obvious. The estimate $\dim_H \mu \leq \underline{\text{ID}}(\mu)$ follows by combining [Fal97, Propositions 10.2–10.3] with [FLR02, Theorem 1.3] and [Fal97, Proposition 10.6]. \square

For more information on dimension theory in Euclidean spaces we refer to [Fal04, Mat95, Rob11].

2.2. Prevalence. In the formulation of our results, the genericity of the considered observables is understood in terms of *prevalence* – a notion introduced by Hunt, Shroer and Yorke in [HSY92], which is regarded to be an analogue of ‘Lebesgue almost sure’ condition in infinite dimensional normed linear spaces.

Definition 2.2. Let V be a normed space. A Borel set $S \subset V$ is called *prevalent* if there exists a Borel measure ν in V , which is positive and finite on some compact set in V , such that for every $v \in V$, the vector $v + e$ belongs to S for ν -almost every $e \in V$. A non-Borel subset of V is prevalent if it contains a prevalent Borel subset.

We will apply this definition to the space $\text{Lip}(X)$ of all Lipschitz functions $h: X \rightarrow \mathbb{R}$ on a compact metric space X , endowed with the Lipschitz norm $\|h\|_{\text{Lip}} = \|h\|_{\infty} + \text{Lip}(h)$, where $\|h\|_{\infty}$ is the supremum norm and $\text{Lip}(h)$ is the Lipschitz constant of h . We will use the following standard condition, which is sufficient for prevalence. Let $\{h_1, \dots, h_m\}$, $m \in \mathbb{N}$, be a finite set of functions in $\text{Lip}(X)$, called the *probe set*. Define $\xi: \mathbb{R}^m \rightarrow \text{Lip}(X)$ by $\xi(\alpha_1, \dots, \alpha_m) = \sum_{j=1}^m \alpha_j h_j$. Then $\nu = \xi_* \text{Leb}$, where Leb is the Lebesgue measure in \mathbb{R}^m , is a Borel measure in $\text{Lip}(X)$, which is positive and finite on the compact set $\xi([0, 1]^m)$. For this measure, the sufficient condition for a set $S \subset \text{Lip}(X)$ to be prevalent is that for every $h \in \text{Lip}(X)$, the function $h + \sum_{j=1}^m \alpha_j h_j$ is in S for Lebesgue almost every $(\alpha_1, \dots, \alpha_m) \in \mathbb{R}^m$. In this case, we say that S is *prevalent* in $\text{Lip}(X)$ with the probe set $\{h_1, \dots, h_m\}$.

For more information on prevalence we refer to [HSY92] and [Rob11, Chapter 5].

2.3. Probabilistic Takens delay embedding theorem. To prove Theorem 1.7, we will use our previous result from [BGŚ20], which we recall below, using the notion of prevalence described in Sect. 2.2. This is a probabilistic version of the Takens delay embedding theorem, asserting that under suitable conditions on k , there is a prevalent set of Lipschitz observables, which give rise to an almost surely injective k -delay coordinate map.

Theorem 2.3 (Probabilistic Takens delay embedding theorem, [BGŚ20, Theorem 4.3 and Remark 4.4]). *Let $X \subset \mathbb{R}^N$ be a compact set, μ a Borel probability measure on X and $T: X \rightarrow X$ an injective Lipschitz map. Take $k > \dim_H \mu$ and assume $\dim_H(\mu|_{\text{Per}_p(T)}) < p$ for $p = 1, \dots, k-1$. Let S be the set of Lipschitz observables $h: X \rightarrow \mathbb{R}$, for which the k -delay coordinate map $\phi_{h,k}$ is injective on a Borel set $X_h \subset X$ of full μ -measure. Then S is prevalent in $\text{Lip}(X)$ with the probe set equal to a linear basis of the space of real polynomials of N variables of degree at most $2k-1$. If μ is additionally T -invariant, then the set X_h for $h \in S$ can be chosen to satisfy $T(X_h) = X_h$.*

2.4. Topological Rokhlin disintegration theorem. A useful tool connecting the probabilistic Takens delay embedding theorem and the SSOY predictability conjecture is the following topological version of the Rokhlin disintegration theorem in compact metric spaces. The Rokhlin disintegration theorem (see e.g. [Roh52]) is a classical result on the existence and almost sure uniqueness of the system of conditional measures. The crucial fact for us is that in the topological setting, the conditional measures can be defined as limits of conditional measures on preimages of shrinking balls, where the convergence holds almost surely, as was proved by Simmons in [Sim12].

In the context of the Rokhlin disintegration theorem, one assumes that the considered measures are *complete*, i.e. every subset of a zero-measure set is measurable. Recall that every finite Borel measure μ on a metric space X has an extension (*completion*) to a complete measure on the σ -algebra of μ -measurable sets, i.e. the smallest σ -algebra containing all Borel sets in X and all subsets of zero μ -measure Borel sets. In other words, every μ -measurable set A can be expressed as $A = B \cup C$, where B is a Borel set and $C \subset D$ for some Borel set D with $\mu(D) = 0$ (see e.g. [Fol99, Theorem 1.19] for the case $X = \mathbb{R}$). Alternatively, this σ -algebra is obtained as a family of sets measurable with respect to the outer measure generated by μ (see e.g. [Fol99, Example 22, p. 32]). Recall also that a function $\psi: X \rightarrow \mathbb{R}$ is called μ -measurable if $\psi^{-1}(B)$ is μ -measurable for every Borel set $B \subset \mathbb{R}$.

Definition 2.4. Let X be a compact metric space and let μ be a complete Borel probability measure on X . Let Y be a separable Riemannian manifold and let $\phi: X \rightarrow Y$ be a Borel map. Set $\nu = \phi_*\mu$ (considered as a complete Borel measure in Y). A family $\{\mu_y : y \in Y\}$ is a *system of conditional measures* of μ with respect to ϕ , if

- (1) for every $y \in Y$, μ_y is a (possibly zero) Borel measure on $\phi^{-1}(\{y\})$,
- (2) for ν -almost every $y \in Y$, μ_y is a Borel probability measure,
- (3) for every μ -measurable set $A \subset X$, the function $Y \ni y \mapsto \mu_y(A)$ is ν -measurable and

$$\mu(A) = \int_Y \mu_y(A) d\nu(y).$$

We say that system of conditional measures $\{\mu_y : y \in Y\}$ is *unique*, if for every family $\{\tilde{\mu}_y : y \in Y\}$ satisfying (1)–(3), we have $\tilde{\mu}_y = \mu_y$ for ν -almost every $y \in Y$.

Theorem 2.5 (Topological Rokhlin disintegration theorem, [Sim12, Theorems 2.1–2.2]). *Let X be a compact metric space and let μ be a Borel probability measure on X . Let Y be a separable Riemannian manifold and let $\phi: X \rightarrow Y$ be a Borel map. Set $\nu = \phi_*\mu$. Then for ν -almost every $y \in \text{supp } \nu$ and $\varepsilon > 0$, the conditional probability measures*

$$\mu_{y,\varepsilon} = \frac{1}{\mu(\phi^{-1}(B(y, \varepsilon)))} \mu|_{\phi^{-1}(B(y,\varepsilon))}$$

converge in weak- topology to a Borel probability measure μ_y as ε tends to 0. Moreover, the collection of measures $\{\mu_y : y \in Y\}$, where we set $\mu_y = 0$ if $y \notin \text{supp } \nu$ or the convergence does not hold, is a unique system of conditional measures of μ with respect to ϕ .*

The proof of the above theorem is based on the differentiation theorem for finite Borel measures, see [Sim12, Theorem 9.1] for details.

3. Proof of the Predictable Embedding Theorem for Lipschitz Maps

In this section we prove the following extended version of Theorem 1.7, which at the same time is an extension of Theorem 2.3 asserting prevalent almost sure predictability.

Theorem 3.1 (Predictable embedding theorem for Lipschitz maps – extended version). *Let $X \subset \mathbb{R}^N$ be a compact set, let μ be a Borel probability measure on X and let $T: X \rightarrow X$ be an injective and Lipschitz map. Take $k > \dim_H \mu$ and assume $\dim_H(\mu|_{\text{Per}_p(T)}) < p$ for $p = 1, \dots, k - 1$. Then there is a set S of Lipschitz observables $h: X \rightarrow \mathbb{R}$, such that S is prevalent in $\text{Lip}(X)$ with the probe set equal to a linear basis of the space of real polynomials of N variables of degree at most $2k - 1$, and for every $h \in S$, the following assertions hold.*

- (a) *There exists a Borel set $X_h \subset X$ of full μ -measure, such that the k -delay coordinate map $\phi_{h,k}$ is injective on X_h .*
- (b) *For every $x \in X_h$, $\lim_{\varepsilon \rightarrow 0} \mu_{\phi_{h,k}(x),\varepsilon} = \delta_x$ in the weak-* topology, where δ_x denotes the Dirac measure at the point x .*
- (c) *$\nu_{h,k}$ -almost every point of \mathbb{R}^k is predictable.*

If μ is additionally T -invariant, then the set X_h for $h \in S$ can be chosen to satisfy $T(X_h) = X_h$.

The main ingredients of the proof of Theorem 3.1 are Theorems 2.3 and 2.5. First, notice that under the assumptions of Theorem 3.1, we can use Theorem 2.5 to show the existence of a system $\{\mu_y : y \in \mathbb{R}^k\}$ of conditional measures of μ with respect to $\phi_{h,k}$, such that for $\nu_{h,k}$ -almost every $y \in \mathbb{R}^k$, μ_y is a Borel probability measure in X satisfying

$$\mu_y = \lim_{\varepsilon \rightarrow 0} \mu_{y,\varepsilon} \tag{3.1}$$

in weak-* topology, where

$$\mu_{y,\varepsilon} = \frac{1}{\mu(\phi_{h,k}^{-1}(B(y, \varepsilon)))} \mu|_{\phi_{h,k}^{-1}(B(y,\varepsilon))}$$

for $\varepsilon > 0$.

The following lemma shows that for $\nu_{h,k}$ -almost every $y \in \mathbb{R}^k$, the prediction error $\sigma(y)$ from Definition 1.1 is equal to the standard deviation of the random variable $\phi_{h,k} \circ T$ with respect to the measure μ_y . Note that the lemma is valid for any continuous (non-necessary Lipschitz) maps T and h .

Lemma 3.2. For $\nu_{h,k}$ -almost every $y \in \mathbb{R}^k$,

$$\sigma(y) = \sqrt{\text{Var}_{\mu_y}(\phi_{h,k} \circ T)},$$

where

$$\text{Var}_{\mu_y}(\phi_{h,k} \circ T) = \int_X \left\| \phi_{h,k} \circ T - \int_X \phi_{h,k} \circ T d\mu_y \right\|^2 d\mu_y.$$

Proof. For simplicity, let us write $\phi = \phi_{h,k}$. Observe first that for $\nu_{h,k}$ -almost every $y \in \mathbb{R}^k$, by (3.1) and the continuity of $\phi \circ T$, we have

$$\chi_\varepsilon(y) = \int_X \phi \circ T d\mu_{y,\varepsilon} \xrightarrow{\varepsilon \rightarrow 0} \chi(y) \tag{3.2}$$

for

$$\chi(y) = \int_X \phi \circ T d\mu_y.$$

Moreover,

$$\begin{aligned} \sigma_\varepsilon^2(y) - \text{Var}_{\mu_y}(\phi \circ T) &= \int_X \|\phi \circ T - \chi_\varepsilon(y)\|^2 d\mu_{y,\varepsilon} - \int_X \|\phi \circ T - \chi(y)\|^2 d\mu_y \\ &= \int_X \|\phi \circ T - \chi_\varepsilon(y)\|^2 d\mu_{y,\varepsilon} - \int_X \|\phi \circ T - \chi(y)\|^2 d\mu_{y,\varepsilon} \end{aligned}$$

$$\begin{aligned}
 & + \int_X \|\phi \circ T - \chi(y)\|^2 d\mu_{y,\varepsilon} - \int_X \|\phi \circ T - \chi(y)\|^2 d\mu_y, \\
 & = I + II.
 \end{aligned}$$

Again by the continuity of $\phi \circ T$, we have $II \xrightarrow{\varepsilon \rightarrow 0} 0$. Furthermore,

$$\begin{aligned}
 |I| & \leq \int_X \left| \|\phi \circ T - \chi_\varepsilon(y)\|^2 - \|\phi \circ T - \chi(y)\|^2 \right| d\mu_{y,\varepsilon} \\
 & = \int_X (\|\phi \circ T - \chi_\varepsilon(y)\| + \|\phi \circ T - \chi(y)\|) \left| \|\phi \circ T - \chi_\varepsilon(y)\| - \|\phi \circ T - \chi(y)\| \right| d\mu_{y,\varepsilon} \\
 & \leq 4\|\phi \circ T\|_\infty \int_X \|\chi_\varepsilon(y) - \chi(y)\| d\mu_{y,\varepsilon} = 4\|\phi \circ T\|_\infty \|\chi_\varepsilon - \chi\|,
 \end{aligned}$$

by the triangle inequality and the fact $\chi_\varepsilon(y) \leq \|\phi \circ T\|_\infty$. The latter quantity converges to zero by (3.2). Therefore, $\sigma_\varepsilon^2(y)$ tends to $\text{Var}_{\mu_y}(\phi \circ T)$ as $\varepsilon \rightarrow 0$, so $\sigma(y) = \sqrt{\text{Var}_{\mu_y}(\phi \circ T)}$. \square

The following corollary is immediate.

Corollary 3.3. *For $\nu_{h,k}$ -almost every $y \in \mathbb{R}^k$, y is predictable if and only if $\phi_{h,k} \circ T$ is constant μ_y -almost surely. In particular, y is predictable provided $\mu_y = \delta_x$ for some $x \in X$.*

By Corollary 3.3, in order to establish almost sure predictability, it is enough to prove the convergence $\lim_{\varepsilon \rightarrow 0} \mu_{\phi_{h,k}(x),\varepsilon} = \delta_x$ for almost every $x \in X$. The idea of the proof of Theorem 3.1 is the following. Theorem 2.3 guarantees that for a prevalent set of observables, the corresponding delay-coordinate map is injective on a set of full μ -measure. On the other hand, Theorem 2.5 assures that the measures $\mu_{\phi(x),\varepsilon}$ are almost surely convergent as $\varepsilon \rightarrow 0$, and the limits form a system of conditional measures of μ with respect to $\phi_{h,k}$. Almost sure injectivity implies that these conditional measures are almost surely Dirac measures, hence indeed $\lim_{\varepsilon \rightarrow 0} \mu_{\phi_{h,k}(x),\varepsilon} = \delta_x$. A detailed proof is presented below.

Proof of Theorem 3.1. By Theorem 2.3, there exists a prevalent set S of Lipschitz observables h , such that for each $h \in S$, the k -delay coordinate map $\phi_{h,k}$ is injective on a Borel set $\tilde{X}_h \subset X$ of full μ -measure. For $h \in S$, let us denote for simplicity $\phi = \phi_{h,k}$ and

$$\tilde{Y}_h = \phi(\tilde{X}_h).$$

Note that \tilde{Y}_h has full $\nu_{h,k}$ -measure. Moreover, \tilde{Y}_h is Borel, as a continuous and injective image of a Borel set, see [Kec95, Theorem 15.1]. Since ϕ is injective on \tilde{X}_h , for every $y \in \tilde{Y}_h$ there exists a unique point $x_y \in \tilde{X}_h$, such that $\phi(x_y) = y$. For $y \in \mathbb{R}^k$ define

$$\tilde{\mu}_y = \begin{cases} \delta_{x_y} & \text{for } y \in \tilde{Y}_h \\ 0 & \text{for } y \in \mathbb{R}^k \setminus \tilde{Y}_h \end{cases}.$$

We check that the collection $\{\tilde{\mu}_y : y \in \mathbb{R}^k\}$ satisfies the conditions (1)–(3) of Definition 2.4. The first two conditions are obvious. To check the third one, take a μ -measurable set $A \subset X$ and note that for $y \in \phi(A \cap \tilde{X}_h)$, we have $y \in \tilde{Y}_h$ and $x_y \in A$, so $\tilde{\mu}_y(A) = \delta_{x_y}(A) = 1$. On the other hand, if $y \in \tilde{Y}_h \setminus \phi(A \cap \tilde{X}_h)$, then $x_y \notin A$, so $\tilde{\mu}_y(A) = \delta_{x_y}(A) = 0$. Since $\tilde{\mu}_y(A) = 0$ for $y \in \mathbb{R}^k \setminus \tilde{Y}_h$, we conclude that for

$$\psi: \mathbb{R}^k \rightarrow \mathbb{R}, \quad \psi(y) = \tilde{\mu}_y(A)$$

we have

$$\psi = \mathbb{1}_{\phi(A \cap \tilde{X}_h)}. \tag{3.3}$$

Hence, to show the $\nu_{h,k}$ -measurability of ψ , it is enough to check that the set $\phi(A \cap \tilde{X}_h)$ is $\nu_{h,k}$ -measurable. To do it, note that since A is μ -measurable, we have $A = B \cup C$, where B is a Borel set and $C \subset D$ for some Borel set D with $\mu(D) = 0$. Hence, $\phi(A \cap \tilde{X}_h) = \phi(B \cap \tilde{X}_h) \cup \phi(C \cap \tilde{X}_h)$. The set $\phi(B \cap \tilde{X}_h)$ is Borel, which again follows from [Kec95, Theorem 15.1], as ϕ is continuous and injective on the Borel set $B \cap \tilde{X}_h$. Similarly, the set $\phi(C \cap \tilde{X}_h)$ is contained in the Borel set $\phi(D \cap \tilde{X}_h)$. Since \tilde{X}_h has full μ -measure, we have

$$\nu_{h,k}(\phi(D \cap \tilde{X}_h)) = \mu(\phi^{-1}(\phi(D \cap \tilde{X}_h))) = \mu(\phi^{-1}(\phi(D \cap \tilde{X}_h)) \cap \tilde{X}_h) = \mu(D) = 0.$$

This yields the $\nu_{h,k}$ -measurability of the set $\phi(A \cap \tilde{X}_h)$ and the function ψ . Moreover, by (3.3),

$$\begin{aligned} \int_Y \tilde{\mu}_y(A) d\nu_{h,k}(y) &= \nu_{h,k}(\phi(A \cap \tilde{X}_h)) \\ &= \mu(\phi^{-1}(\phi(A \cap \tilde{X}_h))) \\ &= \mu(\phi^{-1}(\phi(A \cap \tilde{X}_h)) \cap \tilde{X}_h) = \mu(A). \end{aligned}$$

It follows that $\{\tilde{\mu}_y : y \in \mathbb{R}^k\}$ is a system of conditional measures of μ with respect to ϕ , so by the uniqueness in Theorem 2.5 and (3.1),

$$\tilde{\mu}_y = \mu_y = \lim_{\varepsilon \rightarrow 0} \mu_{y,\varepsilon}$$

for $\nu_{h,k}$ -almost every $y \in \mathbb{R}^k$. Since \tilde{Y}_h is a Borel set of full $\nu_{h,k}$ -measure, we have

$$\mu_y = \lim_{\varepsilon \rightarrow 0} \mu_{y,\varepsilon} = \delta_{x_y} \tag{3.4}$$

for every $y \in Y_h$, where $Y_h \subset \tilde{Y}_h$ and Y_h is a Borel set of full $\nu_{h,k}$ -measure. By Corollary 3.3, this implies that $\nu_{h,k}$ -almost every $y \in \mathbb{R}^k$ is predictable, which proves the assertion (c) in Theorem 3.1.

Define

$$X_h = \phi^{-1}(Y_h) \cap \tilde{X}_h.$$

Then X_h is a Borel full μ -measure subset of X . Since $\phi(X_h) \subset Y_h \subset \tilde{Y}_h$, by (3.4) we have

$$\mu_{\phi(x)} = \lim_{\varepsilon \rightarrow 0} \mu_{\phi(x),\varepsilon} = \delta_{x_{\phi(x)}} = \delta_x$$

for every $x \in X_h$, which shows the assertion (b). Finally, the assertion (a) follows from the fact $X_h \subset \bar{X}_h$.

To end the proof of Theorem 3.1, note that if the measure μ is T -invariant, we can define $X'_h = \bigcap_{n \in \mathbb{Z}} T^n(X_h)$ to obtain a full μ -measure subset of X_h with $T(X'_h) = X'_h$. For details, see the proof of [BGŚ20, Remark 4.4(b)]. \square

Remark 3.4. Similarly as in [BGŚ20], the assumptions $\dim_H(\mu) < k$ and $\dim_H(\mu|_{\text{Per}_p(T)}) < p$ of Theorem 3.1 can be weakened to $\mu \perp \mathcal{H}^k$ and $\mu|_{\text{Per}_p(T)} \perp \mathcal{H}^p$, respectively. Moreover, one can prove a version of Theorem 3.1 for β -Hölder observables $h: X \rightarrow \mathbb{R}$, $\beta \in (0, 1]$. It is enough to take k with $\mathcal{H}^{\beta k}(X) = 0$ and assume that $\mu|_{\text{Per}_p(T)}$ is singular with respect to $\mathcal{H}^{\beta p}$ for $p = 1, \dots, k - 1$, where \mathcal{H}^s is the s -Hausdorff measure. For a precise formulation of required assumptions see [BGŚ20, Theorem 4.3]. As previously, the assumption on periodic points can be omitted if the measure μ is T -invariant and ergodic (see [BGŚ20, Remark 4.4(c)] and its proof).

4. Counterexample to SSOY Predictability Conjecture: Proof of Theorem 1.11

In this section we prove Theorem 1.11, constructing an example of a C^∞ -smooth diffeomorphism T of a compact Riemannian manifold X with an attractor Λ endowed with a natural measure μ , such that $\text{ID}(\mu) < 1$ and for a prevalent set of Lipschitz observables, there is a positive $\nu_{h,1}$ -measure set of non-predictable points. In particular, the set of Lipschitz observables $h: X \rightarrow \mathbb{R}$, for which $\nu_{h,1}$ -almost sure predictability holds for the 1-delay coordinate map $\phi_{h,1}$, is not prevalent. Since the proof is quite involved, we shortly describe the subsequent steps.

In Sect. 4.1 we construct a model for the natural measure μ . First, we prove that for an irrational rotation on a circle $\mathbb{S}^1 \subset \mathbb{R}^N$ endowed with the Lebesgue measure $\text{Leb}_{\mathbb{S}^1}$, the only Lipschitz observables $h: \mathbb{S}^1 \rightarrow \mathbb{R}$ such that the almost sure predictability holds for the 1-delay coordinate map ϕ , are the constant functions. Then we construct a model μ_0 for the natural measure μ , taking $X_0 = \{p_0\} \cup \mathbb{S}^1 \subset \mathbb{R}^N$ for some $p_0 \notin \mathbb{S}^1$ and defining $T_0: X_0 \rightarrow X_0$ as the identity on $\{p_0\}$ and an irrational rotation on \mathbb{S}^1 . Then the measure $\mu_0 = \delta_{p_0}/2 + \text{Leb}_{\mathbb{S}^1}/2$ satisfies $\text{ID}(\mu_0) = 1/2 < 1$, yet the only Lipschitz observables $h: X_0 \rightarrow \mathbb{R}$ yielding almost sure predictability for the 1-delay coordinate maps are the functions constant on \mathbb{S}^1 . The same holds for any extension (X, μ, T) of (X_0, μ_0, T_0) with $X_0 \subset X$, $T|_{X_0} = T_0$ and $\mu = \mu_0$. In particular, the set of Lipschitz observables $h: X \rightarrow \mathbb{R}$ with almost sure predictability for the 1-delay coordinate map, is not prevalent. Moreover, for a prevalent set of Lipschitz observables, the almost sure predictability does not hold (Corollary 4.3).

The main step, performed in Sects. 4.2–4.3 is to realize the model measure μ_0 as a natural measure μ for a smooth diffeomorphism T of a compact Riemannian manifold X . In Sect. 4.2, we construct a C^∞ -diffeomorphism f of the 2-dimensional sphere $\mathbb{S}^2 = \mathbb{R}^2 \cup \{\infty\}$, such that the trajectories of Lebesgue-almost all points of \mathbb{S}^2 spiral towards the invariant unit circle $S = \{(x, y) : x^2 + y^2 = 1\}$, spending most of the time in small neighbourhoods of two fixed points $p, q \in S$ (Proposition 4.12). It follows that the average of the Dirac measures at p and q is a natural measure for f , with the attractor S and basin $\mathbb{S}^2 \setminus \{(0, 0), \infty\}$ (Corollary 4.13). Then, in Sect. 4.3, we take

$$X = \mathbb{S}^2 \times \mathbb{S}^1$$

and define a C^∞ -diffeomorphism $T: X \rightarrow X$ as a skew product of the form

$$T(z, t) = (f(z), h_z(t)), \quad z \in \mathbb{S}^2, t \in \mathbb{S}^1,$$

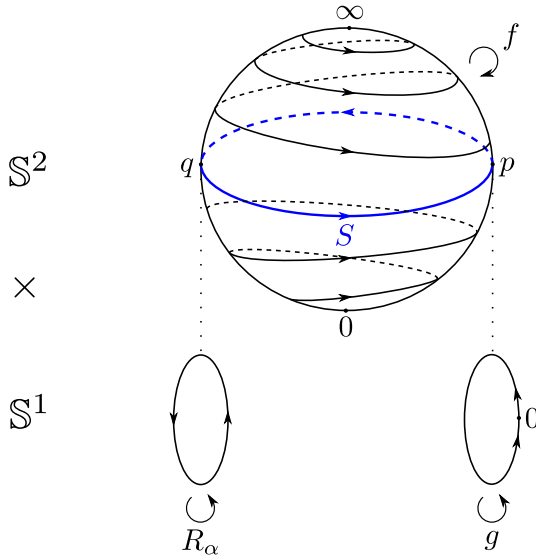


Fig. 1. Schematic view of the map $T : \mathbb{S}^2 \times \mathbb{S}^1 \rightarrow \mathbb{S}^2 \times \mathbb{S}^1$

where h_z are diffeomorphisms of \mathbb{S}^1 depending smoothly on $z \in \mathbb{S}^2$, such that for z in a neighbourhood of p , the map h_z is equal to a map $g : \mathbb{S}^1 \rightarrow \mathbb{S}^1$ with a unique fixed point $0 \in \mathbb{R}/\mathbb{Z} \simeq \mathbb{S}^1$ attracting all points of \mathbb{S}^1 , while for z in a neighbourhood of q , the map h_z is an irrational rotation on \mathbb{S}^1 . See Fig. 1 for a schematic view of the map T .

The map T has an attractor

$$\Lambda = S \times \mathbb{S}^1$$

with the basin $B(\Lambda) = (\mathbb{S}^2 \setminus \{0, \infty\}) \times \mathbb{S}^1$ and natural measure

$$\mu = \frac{1}{2} \delta_{p_0} + \frac{1}{2} \text{Leb}_{\mathbb{S}^1},$$

where $p_0 = (p, 0)$ and $\text{Leb}_{\mathbb{S}^1}$ is the Lebesgue measure on the circle $\{q\} \times \mathbb{S}^1$ (Theorem 4.14). Since the measure μ is equal to the model measure μ_0 , the conclusion follows.

4.1. Model measure. Consider a circle $\mathbb{S}^1 \subset \mathbb{R}^N$ (by a circle we mean an image of $\{(x, y) \in \mathbb{R}^2 : x^2 + y^2 = 1\}$ by an affine similarity transformation) with the normalized Lebesgue (1-Hausdorff) measure $\text{Leb}_{\mathbb{S}^1}$ and a rotation $R_\alpha : \mathbb{S}^1 \rightarrow \mathbb{S}^1$ by an angle α . We use here an additive notation, i.e. for an angle coordinate $t \in \mathbb{R}/\mathbb{Z} \simeq \mathbb{S}^1$ we write $R_\alpha(t) = t + \alpha \pmod{1}$. We assume $\alpha \in \mathbb{R} \setminus \mathbb{Q}$. By $d(\cdot, \cdot)$ we denote the standard rotation-invariant metric on \mathbb{S}^1 .

For the system $(\mathbb{S}^1, \text{Leb}_{\mathbb{S}^1}, R_\alpha)$ we consider Lipschitz observables $h : \mathbb{S}^1 \rightarrow \mathbb{R}$ and the corresponding 1-delay coordinate maps $\phi : \mathbb{S}^1 \rightarrow \mathbb{R}$. Note that 1-delay coordinate maps are equal to the observables, i.e. $\phi = h$.

Proposition 4.1. *Suppose that for a Lipschitz function $h : \mathbb{S}^1 \rightarrow \mathbb{R}$, ν -almost every $y \in \mathbb{R}$ is predictable for the 1-delay coordinate map $\phi = h$, where $\nu = \phi_* \text{Leb}_{\mathbb{S}^1}$. Then h is constant.*

Proof. Take h as in the proposition. The proof that h is constant is divided into four parts, described by the following claims.

Claim 1. *There exists a set $B \subset \mathbb{S}^1$ of full $\text{Leb}_{\mathbb{S}^1}$ -measure, with the following property: if $t_1, t_2 \in B$ and $h(t_1) = h(t_2)$, then $h(R_\alpha^n t_1) = h(R_\alpha^n t_2)$ for every $n \geq 0$.*

For the proof of the above claim, consider the system $\{\mu_y : y \in \mathbb{R}\}$ of conditional measures of $\text{Leb}_{\mathbb{S}^1}$ with respect to $\phi = h$, given by Theorem 2.5. Let

$$A = \left\{ t \in \mathbb{S}^1 : h(R_\alpha t) = \int h \circ R_\alpha d\mu_{h(t)} \right\}.$$

It follows from Theorem 2.5 that the map $y \mapsto \int h \circ R_\alpha d\mu_y$ is ν -measurable, hence $t \mapsto \int h \circ R_\alpha d\mu_{h(t)}$ is $\text{Leb}_{\mathbb{S}^1}$ -measurable. Consequently, A is a $\text{Leb}_{\mathbb{S}^1}$ -measurable set. By Theorem 2.5,

$$\text{Leb}_{\mathbb{S}^1}(A) = \int_{\mathbb{R}} \mu_y(A) d\nu(y) \tag{4.1}$$

and

$$\mu_y(A) = \mu_y(A \cap \{h = y\}) = \mu_y\left(\left\{ t \in \mathbb{S}^1 : h(t) = y \text{ and } h(R_\alpha t) = \int h \circ R_\alpha d\mu_y \right\}\right).$$

Since ν -almost every $y \in \mathbb{R}$ is predictable, Lemma 3.2 implies that the function $h \circ R_\alpha$ is constant μ_y -almost surely for ν -almost every $y \in \mathbb{R}$, hence $\mu_y(A) = 1$ for ν -almost every $y \in \mathbb{R}$. Therefore, (4.1) gives $\text{Leb}_{\mathbb{S}^1}(A) = 1$.

Let

$$B = \bigcap_{n=0}^{\infty} R_\alpha^{-n}(A).$$

Then B has full $\text{Leb}_{\mathbb{S}^1}$ -measure. Moreover, the definition of A implies that if $t_1, t_2 \in A$ and $h(t_1) = h(t_2)$, then $h(R_\alpha t_1) = h(R_\alpha t_2)$. Therefore, if $t_1, t_2 \in B$ and $h(t_1) = h(t_2)$, then $h(R_\alpha^n t_1) = h(R_\alpha^n t_2)$ for every $n \geq 0$.

Claim 2. *If $t_1, t_2 \in B$ and $h(t_1) = h(t_2)$, then $h(t_1 + s) = h(t_2 + s)$ for every $s \in \mathbb{S}^1$.*

In order to prove the claim, assume that $t_1, t_2 \in B$ and $h(t_1) = h(t_2)$. Fix $s \in \mathbb{S}^1$. Since $\alpha \notin \mathbb{Q}$, every orbit under R_α is dense in \mathbb{S}^1 , so there exists a sequence $n_k \rightarrow \infty$ with $R_\alpha^{n_k} t_1 \rightarrow t_1 + s$ as $k \rightarrow \infty$. Then $R_\alpha^{n_k} t_2 \rightarrow t_2 + s$. As $t_1, t_2 \in B$ and $h(t_1) = h(t_2)$, by Claim 1 we have $h(R_\alpha^{n_k} t_1) = h(R_\alpha^{n_k} t_2)$, hence the continuity of h gives $h(t_1 + s) = h(t_2 + s)$.

Claim 3. *For every $\varepsilon > 0$, there exist $t_1, t_2 \in B$ such that $0 < d(t_1, t_2) < \varepsilon$ and $h(t_1) = h(t_2)$.*

To prove Claim 3, note first that it holds trivially if the set $h^{-1}(\{\inf h\})$ has non-empty interior. Otherwise, fix a small $\varepsilon > 0$ and take $t_0 \in \mathbb{S}^1$ such that $h(t_0) = \inf h$. Then by the continuity of h , there exist disjoint open arcs $I, J \subset \mathbb{S}^1$ of length smaller than $\varepsilon/2$, such that $\overline{I} \cap \overline{J} = \{t_0\}$ and their images $h(I), h(J)$ are intervals of positive length with $\overline{h(I)} = \overline{h(J)} = K$ for some closed, non-degenerate interval $K \subset \mathbb{R}$. As B is of full $\text{Leb}_{\mathbb{S}^1}$ -measure and h is Lipschitz, $h(I \cap B)$ and $h(J \cap B)$ both have full Lebesgue measure in K , hence $h(I \cap B) \cap h(J \cap B) \neq \emptyset$. This proves the claim.

Claim 4. *The map h is constant.*

For the proof of Claim 4, fix a small $\delta > 0$. As h is uniformly continuous, there exists $\varepsilon > 0$ such that $|h(t) - h(t')| < \delta$ whenever $d(t, t') < \varepsilon$. According to Claim 3, there exist $t_1, t_2 \in B$ such that $0 < d(t_1, t_2) < \varepsilon$ and $h(t_1) = h(t_2)$. Let $\beta = t_2 - t_1 \bmod 1$ and note that $\beta \neq 0, |\beta| < \varepsilon$. Applying inductively Claim 2 to t_1, t_2 with $s = \beta, \dots, (n - 1)\beta \bmod 1$, for $n \in \mathbb{N}$, we obtain $h(t_1) = h(t_1 + \beta \bmod 1) = \dots = h(t_1 + n\beta \bmod 1)$. Again by Claim 2, we arrive at $h(0) = h(n\beta \bmod 1)$ for $n \in \mathbb{N}$.

Take $t \in \mathbb{S}^1$. As $|\beta| < \varepsilon$, for every $t \in \mathbb{S}^1$ there exists $n \in \mathbb{N}$ such that $d(t, n\beta \bmod 1) < \varepsilon$. For such n we have $|h(t) - h(0)| = |h(t) - h(n\beta \bmod 1)| < \delta$. As δ was arbitrary, we have $h(t) = h(0)$. Therefore, h is constant. \square

Remark 4.2. In [BGS20, Example 3.5] it is shown that there does not exist a Lipschitz map $h: \mathbb{S}^1 \rightarrow \mathbb{R}$ which is injective on a set of full $\text{Leb}_{\mathbb{S}^1}$ -measure. However, it may still happen that for certain Lipschitz transformations $T: \mathbb{S}^1 \rightarrow \mathbb{S}^1$ almost sure predictability holds for every h , e.g. if T is the identity.

Corollary 4.3. *Let $X \subset \mathbb{R}^N$ be a compact set with a Borel probability measure μ and let $T: X \rightarrow X$ be an injective Lipschitz map, such that*

$$(\text{supp } \mu, \mu, T|_{\text{supp } \mu}) = (X_0, \mu_0, T_0),$$

where $X_0 = \{p_0\} \cup \mathbb{S}^1$ for a circle $\mathbb{S}^1 \subset \mathbb{R}^N$ and $p_0 \in \mathbb{R}^N \setminus \mathbb{S}^1$,

$$\mu_0 = \frac{1}{2} \delta_{p_0} + \frac{1}{2} \text{Leb}_{\mathbb{S}^1},$$

and $T_0: X_0 \rightarrow X_0$, such that $T_0(p_0) = p_0$ and T_0 is an irrational rotation R_α on \mathbb{S}^1 . Set $\nu = \phi_* \mu$. Then $\text{ID}(\mu) = 1/2$ and the only Lipschitz observables $h: X \rightarrow \mathbb{R}$, such that ν -almost every $y \in \mathbb{R}^k$ is predictable for the 1-delay coordinate map $\phi = h$, are the ones constant on \mathbb{S}^1 . Consequently, for a prevalent set of Lipschitz observables, there is a positive ν -measure set of non-predictable points. In particular, the set of Lipschitz observables $h: X \rightarrow \mathbb{R}$ for which ν -almost every point of \mathbb{R} is predictable, is not prevalent.

Proof. The fact $\text{ID}(\mu) = \text{ID}(\mu_0) = 1/2$ follows from the definition of the information by a direct checking. The assertion that only observables constant on \mathbb{S}^1 give almost sure predictability is an immediate consequence of Proposition 4.1. Consider now the space $\text{Lip}(X)$ of all Lipschitz observables $T: X \rightarrow X$, with the Lipschitz norm $\|h\|_{\text{Lip}}$ (see Sect. 2.2), and let $Z \subset \text{Lip}(X)$ be the set of Lipschitz observables which are constant on \mathbb{S}^1 . Note first that any prevalent set is dense (see [Rob11, Sect. 5.1]), while Z is not dense in $\text{Lip}(X)$ in the supremum norm (hence also in the Lipschitz norm). Therefore, Z is not prevalent in $\text{Lip}(X)$. In fact, we can prove more, showing that $\text{Lip}(X) \setminus Z$ is prevalent (note that a subset of the complement of a prevalent set cannot be prevalent, as the intersection of two prevalent sets is prevalent, see [HSY92]).

In order to prove prevalence of $\text{Lip}(X) \setminus Z$, we can assume that the circle $\mathbb{S}^1 \subset X$ is of the form $\mathbb{S}^1 = \{(x_1, \dots, x_N) \in \mathbb{R}^N : x_1^2 + x_2^2 = 1, x_3 = 0, \dots, x_N = 0\}$. Indeed, an affine change of coordinates in \mathbb{R}^N transforming the circle in X to the circle $\{(x_1, \dots, x_N) \in \mathbb{R}^N : x_1^2 + x_2^2 = 1, x_3 = 0, \dots, x_N = 0\}$ induces a linear isomorphism between the corresponding spaces of Lipschitz observables. Like in Theorem 3.1, we show the prevalence of $\text{Lip}(X) \setminus Z$ with the probe set equal to a linear basis of the space of real polynomials of N variables of degree at most 1. In other words, we should check

that for any $h \in \text{Lip}(X)$, we have $h + \alpha_0 + \alpha_1 h_1 + \dots + \alpha_N h_N \notin Z$ for Lebesgue-almost every $\alpha = (\alpha_0, \dots, \alpha_N) \in \mathbb{R}^{N+1}$, where $h_j(x_1, \dots, x_N) = x_j, j = 1, \dots, N$. Let e_1, \dots, e_N be the standard basis of \mathbb{R}^N . If $h + \alpha_0 + \alpha_1 h_1 + \dots + \alpha_N h_N \in Z$, then evaluating at $e_1, e_2 \in \mathbb{S}^1$ gives

$$h(e_1) + \alpha_0 + \alpha_1 = h(e_2) + \alpha_0 + \alpha_2.$$

Therefore $\alpha_1 = \alpha_2 + h(e_2) - h(e_1)$, so α belongs to an affine subspace of \mathbb{R}^{N+1} of codimension one. It follows that given $h \in \text{Lip}(X)$, we have $h + \alpha_0 + \alpha_1 h_1 + \dots + \alpha_N h_N \in Z$ for $(\alpha_0, \dots, \alpha_N)$ in a set of zero Lebesgue measure in \mathbb{R}^{N+1} , which ends the proof. \square

4.2. Construction of the diffeomorphism $f: \mathbb{S}^2 \rightarrow \mathbb{S}^2$. In this subsection we construct a smooth diffeomorphism f of $\mathbb{S}^2 \simeq \mathbb{R}^2 \cup \{\infty\}$ with the invariant unit circle S containing two fixed points p, q , such that the trajectories of all points in $\mathbb{R}^2 \setminus \{0, 0\}$ spiral towards the invariant unit circle S , spending most of the time in small neighbourhoods of p and q .

We consider points $(x, y) \in \mathbb{R}^2$ in polar coordinates, i.e. $x = r \cos \varphi, y = r \sin \varphi$ for $r \in [0, +\infty), \varphi \in \mathbb{R}$. Let

$$f(r \cos \varphi, r \sin \varphi) = (R(r) \cos \Phi(r, \varphi), R(r) \sin \Phi(r, \varphi))$$

for

$$R(r) = r + \varepsilon \frac{r(1-r)^3}{1+r^4}, \quad \Phi(r, \varphi) = \varphi + \varepsilon \theta(\varphi) + (1-r)^2 \eta(r),$$

where $\varepsilon > 0$ is a small constant, $\theta: \mathbb{R} \rightarrow [0, +\infty)$ is a π -periodic C^∞ -function such that $\theta(\varphi) = \varphi^2$ for $\varphi \in (-\pi/4, \pi/4)$ and θ has no zeroes except for $k\pi, k \in \mathbb{Z}$, while $\eta: [0, +\infty) \rightarrow [0, +\infty)$ is a C^∞ -function such that $\eta|_{[\frac{1}{2}, \frac{3}{2}]} \equiv 1, \eta > 0$ on $(0, \infty)$ and $\lim_{r \rightarrow 0^+} (1-r)^2 \eta(r) = \lim_{r \rightarrow +\infty} (1-r)^2 \eta(r) = 0$ (the role of η is to ensure that f extends to a C^∞ -diffeomorphism of the sphere). The following two lemmas are elementary.

Lemma 4.4. *For sufficiently small $\varepsilon > 0$, the function R has the following properties.*

- (a) R is an increasing homeomorphism of $[0, +\infty)$.
- (b) $R(0) = 0, R(r) > r$ for $r \in (0, 1), R(1) = 1$ and $R(r) < r$ for $r \in (1, +\infty)$.
- (c) Near $r = 1, R$ has the Taylor expansion $R(r) = 1 + r - 1 - \frac{\varepsilon}{2}(r-1)^3 + \dots$.

Lemma 4.5. *For sufficiently small $\varepsilon > 0$, the function Φ has the following properties.*

- (a) $\Phi(r, \varphi) > \varphi$ for $r \in ((0, 1) \cup (1, +\infty))$.
- (b) For given $r \in (0, +\infty)$, the function $\varphi \mapsto \Phi(r, \varphi)$ is strictly increasing.
- (c) For the function $\varphi \mapsto \Phi(1, \varphi) \pmod{2\pi}$, the points $0, \pi$ are the unique fixed points and the intervals $(0, \pi), (\pi, 2\pi)$ are invariant.

Let

$$\mathbb{B} = \{(x, y) \in \mathbb{R}^2 : \|(x, y)\| < 1\}, \quad S = \{(x, y) \in \mathbb{R}^2 : \|(x, y)\| = 1\},$$

where $\|\cdot\|$ denotes the Euclidean norm. For sufficiently small ε , the function f defines a C^∞ -diffeomorphism of \mathbb{R}^2 , such that the unit disc \mathbb{B} , the unit circle S and the complement of $\overline{\mathbb{B}}$ are f -invariant. Compactifying \mathbb{R}^2 to the Riemann sphere $\mathbb{S}^2 \simeq \mathbb{R}^2 \cup \{\infty\}$ and

putting $f(\infty) = \infty$, we extend f to a C^∞ -diffeomorphism of \mathbb{S}^2 with fixed points at $(0, 0)$ and ∞ . Another two fixed points,

$$p = (1, 0), \quad q = (-1, 0),$$

corresponding to the fixed points described in Lemma 4.5(c), are located in the unit circle S .

Now we analyse the behaviour of the orbits of points $(x, y) \in \mathbb{S}^2$ under f . By Lemma 4.5, if $(x, y) = (\cos \varphi_0, \sin \varphi_0) \in S$ for some $\varphi_0 \in \mathbb{R}$, then $f^n(x, y)$ tends to p (resp. to q) as $n \rightarrow \infty$ for $\varphi_0 \in (-\pi, 0] \bmod 2\pi$ (resp. $\varphi_0 \in (0, \pi] \bmod 2\pi$). Suppose now $(x, y) \in \mathbb{S}^2 \setminus S$. Recall that the points $(0, 0)$ and ∞ are fixed, so we can assume $(x, y) \in \mathbb{R}^2 \setminus (S \cup \{(0, 0)\})$. Then

$$(x, y) = (r_0 \cos \varphi_0, r_0 \sin \varphi_0)$$

for $r_0 \in \mathbb{R} \setminus \{1\}$, $\varphi_0 \in \mathbb{R}$. The goal of this subsection is to prove

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=0}^{N-1} \delta_{f^n(x,y)} = \frac{1}{2} \delta_p + \frac{1}{2} \delta_q$$

in the sense of weak-* convergence (see Corollary 4.13). To this aim, we find the asymptotics of the subsequent times spent by the iterates of (x, y) in small neighbourhoods of the points p and q . We will make calculations only for the case

$$r_0 \in (0, 1),$$

since the functions R, Φ are defined such that the behaviour of the trajectories in the case of points $r_0 > 1$ is symmetric (see Remark 4.11). From now on, we fix the initial point $(x, y) = (r_0 \cos \varphi_0, r_0 \sin \varphi_0)$ with $r_0 \in (0, 1)$ and allow all the constants appearing below to depend on this point. For $n \in \mathbb{N}$ let

$$r_n = R^n(r_0)$$

and define inductively

$$\varphi_{n+1} = \Phi(r_n, \varphi_n).$$

Then

$$f^n(r_0 \cos \varphi_0, r_0 \sin \varphi_0) = (r_n \cos \varphi_n, r_n \sin \varphi_n).$$

For convenience, set

$$\rho_n = 1 - r_n$$

and note that by Lemma 4.4, ρ_n decreases to 0 as $n \rightarrow \infty$.

Lemma 4.6. *We have*

$$\rho_n = \frac{a + o(1)}{\sqrt{n}}$$

as $n \rightarrow \infty$ for some $a > 0$. Moreover, for every $0 \leq k \leq n$,

$$\frac{k}{cn^{3/2}} \leq \rho_n - \rho_{n+k} \leq \frac{ck}{n^{3/2}},$$

where $c > 0$ is independent of n and k .

Proof. By Lemma 4.4, we have $\rho_n \searrow 0^+$ as $n \rightarrow \infty$ and

$$\rho_{n+1} = \rho_n - \frac{\varepsilon}{2} \rho_n^3 + \dots$$

for ρ_n close to 0. Hence, the first assertion follows from the standard analysis of the behaviour of an analytic map near a parabolic fixed point, see e.g. [Mil06, Lemma 10.1]. To check the second one, note that there exists a univalent holomorphic map $\psi : V \rightarrow \mathbb{C}$ (Fatou coordinate) on a domain $V \subset \mathbb{C}$ containing ρ_n for large n , such that $\psi(V)$ contains a half-plane $\{z \in \mathbb{C} : \operatorname{Re}(z) > c_0\}$ for some $c_0 \in \mathbb{R}$ and

$$\psi(\rho_{n+1}) = \psi(\rho_n) + 1$$

(see e.g. [Mil06, Theorem 10.9]). Let

$$z_n = \psi(\rho_n)$$

for large n and take n_0 with $\operatorname{Re}(z_{n_0}) > c_0$. Then ψ^{-1} is defined on

$$D = \{z \in \mathbb{C} : |z - z_{n+k}| < n + k - n_0\}$$

for large n , and $z_{\lfloor n/2 \rfloor}, z_n \in D'$ for

$$D' = \{z \in \mathbb{C} : |z - z_{n+k}| \leq n + k - \lfloor n/2 \rfloor\}.$$

Since $k \leq n$, the ratio of the radius of D' to the radius of D is at most $\frac{(3/2)n+1}{2n-n_0}$, which tends to $3/4$ as $n \rightarrow \infty$. Moreover,

$$\frac{|z_{n+k} - z_n|}{|z_n - z_{\lfloor n/2 \rfloor}|} = \frac{k}{n - \lfloor n/2 \rfloor}$$

Therefore, by the Koebe distortion theorem (see e.g. [CG93, Theorem 1.6]),

$$\frac{1}{c} \frac{k}{n} < \frac{\rho_n - \rho_{n+k}}{\rho_{\lfloor n/2 \rfloor} - \rho_n} < c \frac{k}{n}$$

for some constant $c > 0$. Since $\sqrt{n}(\rho_{\lfloor n/2 \rfloor} - \rho_n) \rightarrow \sqrt{2} - 1$ as $n \rightarrow \infty$ by the first assertion of the lemma, this ends the proof. \square

Convention. Within subsequent calculations, we will $a_n \asymp b_n$ for sequences a_n, b_n , if $\frac{1}{c} < \frac{a_n}{b_n} < c$, where $c > 0$ is independent of n .

Lemma 4.7. *Suppose*

$$x_{n+1} = x_n + ax_n^2$$

for $n \in \mathbb{Z}$ and some $a > 0$. Then for given $x_0 < 0$ (resp. $x_0 > 0$) sufficiently close to 0, we have

$$x_n \asymp -\frac{1}{n} \quad \left(\text{resp. } x_{-n} \asymp \frac{1}{n} \right)$$

for $n \in \mathbb{N}$.

Proof. Follows directly from [Mil06, Lemma 10.1]. \square

By Lemmas 4.4–4.6, the trajectory of (x, y) approaches the unit circle S , spiralling an infinite number of times near S and slowing down near the fixed points p and q . In fact, the definitions of the functions R, Φ easily imply that p and q are in the limit set of the trajectory. In particular, for a fixed $\delta > 0$ (which is small enough to satisfy several conditions, specified later), the trajectory visits infinitely number of times the δ -neighbourhoods of p and q , defined respectively by

$$\begin{aligned} U_p &= \{(r \cos \varphi, r \sin \varphi) : r \in (1 - \delta, 1 + \delta), \varphi \in (-\delta, \delta)\}, \\ U_q &= \{(r \cos \varphi, r \sin \varphi) : r \in (1 - \delta, 1 + \delta), \varphi \in (\pi - \delta, \pi + \delta)\}. \end{aligned} \tag{4.2}$$

Hence, for $i \in \mathbb{N}$ we can define $N_{p,i}$ (resp. $N_{q,i}$) to be the time spent by the trajectory during its i -th visit in U_p (resp. U_q). More precisely, set $n_{p,0}^+ = 0$ and define inductively

$$\begin{aligned} n_{p,i}^- &= \min\{n \geq n_{p,i-1}^+ : (r_n \cos \varphi_n, r_n \sin \varphi_n) \in U_p\}, \\ n_{p,i}^+ &= \min\{n \geq n_{p,i}^- : (r_n \cos \varphi_n, r_n \sin \varphi_n) \notin U_p\}, \\ N_{p,i} &= n_{p,i}^+ - n_{p,i}^- \end{aligned}$$

for $i \geq 1$. Define $n_{q,i}^-, n_{q,i}^+, N_{q,i}$ analogously. By Lemmas 4.4 and 4.5, if $\delta > 0$ is chosen small enough, then

$$0 < n_{p,1}^- < n_{p,1}^+ < n_{q,1}^- < n_{q,1}^+ < \dots < n_{p,i}^- < n_{p,i}^+ < n_{q,i}^- < n_{q,i}^+ < \dots \tag{4.3}$$

or

$$0 < n_{q,1}^- < n_{q,1}^+ < n_{p,1}^- < n_{p,1}^+ < \dots < n_{q,i}^- < n_{q,i}^+ < n_{p,i}^- < n_{p,i}^+ < \dots,$$

depending on the position of the point (x, y) . To simplify notation, we assume that (4.3) holds. Again by Lemmas 4.4 and 4.5, we obtain the following.

Lemma 4.8. *We have*

$$n_{q,i}^- - n_{p,i}^+, n_{p,i+1}^- - n_{q,i}^+ < N_0$$

for some constant $N_0 > 0$. In other words, the times spent by the trajectory of (x, y) between consecutive visits in $U_p \cup U_q$ remain uniformly bounded.

Now we estimate the times spent by the trajectory during its stay in U_p and U_q .

Lemma 4.9.

$$N_{p,i} \asymp N_{q,i} \asymp i.$$

Proof. We prove the lemma by induction. Obviously, we can assume that i is large. Suppose, by induction,

$$\frac{j}{C} \leq N_{p,j} \leq Cj, \quad \frac{j}{C} \leq N_{q,j} \leq Cj \quad \text{for } j = 1, \dots, i - 1 \tag{4.4}$$

for a large constant $C > 1$ (to be specified later). First, we estimate $N_{p,i}$. By Lemma 4.8,

$$\frac{i^2}{c_1 C} \leq n_{p,i}^- \leq c_1 C i^2 \tag{4.5}$$

for some $c_1 > 0$ (we denote by c_1, c_2, \dots constants independent of C .) Obviously, we can assume $\varphi_{n_{p,i}^-} \in [-\pi, \pi)$. Then, since δ is small and i is large, we have

$$-\frac{\pi}{4} < -\delta < \varphi_{n_{p,i}^-} < 0.$$

Note that $\rho_{n_{p,i}^-} < \delta$ and the sequence ρ_n is decreasing, so

$$N_{p,i} = \min\{n \geq n_{p,i}^- : \varphi_n \geq \delta\} - n_{p,i}^-.$$

Recall that if $\varphi_n \in (-\pi/4, \pi/4)$ (in particular, if $n \in [n_{p,i}^-, n_{p,i}^+)$), then

$$\varphi_{n+1} = \varphi_n + \varepsilon\varphi_n^2 + \rho_n^2. \tag{4.6}$$

Let

$$\rho_i^- = \frac{1}{C^{2/3}i}, \quad \rho_i^+ = \frac{C^{2/3}}{i}.$$

To estimate the behaviour of the sequence φ_n for $n \geq n_{p,i}^-$, we will compare it with the sequences φ_n^+, φ_n^- for $n \geq n_{p,i}^-$, given by

$$\varphi_{n_{p,i}^\pm}^\pm = \varphi_{n_{p,i}^\pm}, \quad \varphi_{n+1}^\pm = \varphi_n^\pm + \varepsilon(\varphi_n^\pm)^2 + (\rho_i^\pm)^2. \tag{4.7}$$

First, we will analyse the behaviour of the sequences φ_n^\pm and then show that they provide upper and lower bounds for φ_n . By definition, $\varphi_{n_{p,i}^\pm}^\pm \in (-\delta, 0)$ and φ_n^\pm increases to infinity as $n \rightarrow \infty$. Hence, we can define

$$N_i^\pm = \min\{n \geq n_{p,i+1}^- : \varphi_n^\pm \geq \delta\} - n_{p,i}^-.$$

to be the time which the sequence φ_n^\pm spends in $(-\delta, \delta)$. Since $\rho_i^- < \rho_i^+$, we have $\varphi_n^- \leq \varphi_n^+$ and $N^+ \leq N^-$. Set

$$k_1^\pm = \min \left\{ n \in [n_{p,i}^-, n_{p,i}^- + N_i^\pm] : \varphi_n^\pm > -\frac{\rho_i^\pm}{\sqrt{\varepsilon}} \right\},$$

$$k_2^\pm = \min \left\{ n \in [k_1^\pm, n_{p,i}^- + N_i^\pm] : \varphi_n^\pm > \frac{\rho_i^\pm}{\sqrt{\varepsilon}} \right\}.$$

Note that for $n \in [n_{p,i}^-, k_1^\pm) \cup [k_2^\pm, N_i^\pm + n_{p,i}^-)$ we have $\varepsilon(\varphi_n^\pm)^2 \geq (\rho_i^\pm)^2$, so

$$\varphi_n^\pm + \varepsilon(\varphi_n^\pm)^2 \leq \varphi_{n+1}^\pm \leq \varphi_n^\pm + 2\varepsilon(\varphi_n^\pm)^2.$$

Hence, by Lemma 4.7,

$$k_1^\pm - n_{p,i}^- \asymp N_i^\pm + n_{p,i}^- - k_2^\pm \asymp \frac{1}{\rho_i^\pm}.$$

On the other hand, for $n \in [k_1^\pm, k_2^\pm)$ we have $\varepsilon(\varphi_n^\pm)^2 \leq (\rho_i^\pm)^2$, so

$$\varphi_n^\pm + (\rho_i^\pm)^2 \leq \varphi_{n+1}^\pm \leq \varphi_n^\pm + 2(\rho_i^\pm)^2,$$

which implies

$$k_2^\pm - k_1^\pm \asymp \frac{1}{\rho_i^\pm}.$$

Hence,

$$\frac{i}{c_2 C^{2/3}} = \frac{1}{c_2 \rho_i^\pm} \leq N_i^+ \leq N_i^- \leq \frac{c_2}{\rho_i^-} = c_2 C^{2/3} i$$

for some $c_2 > 0$. If C is chosen sufficiently large, then this yields

$$\frac{i}{C} \leq N_i^+ \leq N_i^- \leq Ci. \tag{4.8}$$

Now we show by induction that

$$\varphi_n^- \leq \varphi_n \leq \varphi_n^+ \tag{4.9}$$

for $n \in [n_{p,i}^-, n_{p,i}^- + \min(N_{p,i}, N_i^-)]$. To do it, note that for $n = n_{p,i}^-$ we have equalities in (4.9). Suppose, by induction, that (4.9) is satisfied for some $n \in [n_{p,i}^-, n_{p,i}^- + \min(N_{p,i}, N_i^-)]$. Then by (4.6) and (4.7),

$$\varphi_{n+1} - \varphi_{n+1}^\pm = (\varphi_n - \varphi_n^\pm)(1 + \varepsilon(\varphi_n + \varphi_n^\pm)) + \rho_n^2 - (\rho_i^\pm)^2,$$

where $1 + \varepsilon(\varphi_n + \varphi_n^\pm) > 1 - 2\varepsilon\delta > 0$. Moreover, by Lemma 4.6, (4.5) and (4.8), there exists a constant $c_3 > 0$, such that

$$\frac{1}{c_3 \sqrt{C} i} \leq \rho_n \leq \frac{c_3 \sqrt{C}}{i},$$

which gives

$$\rho_i^- \leq \rho_n \leq \rho_i^+,$$

provided C is chosen sufficiently large. Therefore, the sign of $\varphi_{n+1} - \varphi_{n+1}^\pm$ is the same as the one of $\varphi_n - \varphi_n^\pm$, which provides the induction step and proves (4.9).

Using (4.9), we can show

$$N_i^+ \leq N_{p,i} \leq N_i^-. \tag{4.10}$$

Indeed, if $N_{p,i} > N_i^-$, then by (4.9),

$$\delta \leq \varphi_{n_{p,i}^- + N_i^-}^- \leq \varphi_{n_{p,i}^- + N_i^-},$$

so $n_{p,i}^+ \leq n_{p,i}^- + N_i^-$, which is a contradiction. Hence, $N_{p,i} \leq N_i^-$, and then (4.9) gives

$$\delta \leq \varphi_{n_{p,i}^+}^+ \leq \varphi_{n_{p,i}^+}^+,$$

which implies (4.10). By (4.8) and (4.10),

$$\frac{i}{C} \leq N_{p,i} \leq Ci,$$

which completes the inductive step started in (4.4) and shows $N_{p,i} \asymp i$.

To show $N_{q,i} \asymp i$, note that if $\varphi_n \in (3\pi/4, 5\pi/4)$, then for $\tilde{\varphi}_n = \varphi_n - \pi$ we have

$$\tilde{\varphi}_{n+1} = \tilde{\varphi}_n + \varepsilon \tilde{\varphi}_n^2 + \rho_n^2.$$

Moreover, by the proved assertion $N_{p,i} \asymp i$ and Lemmas 4.6 and 4.8, we have $n_{q,i}^- \asymp n_{p,i}^-$ and $\rho_{n_{q,i}^-} \asymp \rho_{n_{p,i}^-}$. Using this, one can show $N_{q,i} \asymp i$ by repeating the proof in the case of $N_{p,i}$. \square

A more accurate comparison of $N_{p,i}$ and $N_{q,i}$ is presented below.

Lemma 4.10. *There exists $M > 0$ such that*

$$|N_{p,i} - N_{q,i}| < M$$

for all $i \geq 1$.

Proof. Take a large $i \in \mathbb{N}$. Let

$$(\eta_n, \psi_n) = f^n(r_{n_{p,i}^-}, \varphi_{n_{p,i}^-}), \quad (\tilde{\eta}_n, \tilde{\psi}_n) = f^n(r_{n_{q,i}^-}, \varphi_{n_{q,i}^-} - \pi)$$

and

$$\sigma_n = 1 - \eta_n = \rho_{n+n_{p,i}^-}, \quad \tilde{\sigma}_n = 1 - \tilde{\eta}_n = \rho_{n+n_{q,i}^-}$$

for $n \geq 0$. Subtracting multiplicities of 2π , we can assume $\psi_0, \tilde{\psi}_0 \in [-\pi, \pi)$, so in fact

$$-\delta < \psi_0, \tilde{\psi}_0 < 0.$$

By definition,

$$\psi_{n+1} = \psi_n + \varepsilon \psi_n^2 + \sigma_n^2, \quad \tilde{\psi}_{n+1} = \tilde{\psi}_n + \varepsilon \tilde{\psi}_n^2 + \tilde{\sigma}_n^2 \tag{4.11}$$

as long as $\psi_n, \tilde{\psi}_n < \pi/4$. It follows that

$$N_{p,i} = \min\{n \geq 0 : \psi_n \geq \delta\}, \quad N_{q,i} = \min\{n \geq 0 : \tilde{\psi}_n \geq \delta\}.$$

Note that (4.11) holds for $n \leq \min(N_{p,i}, N_{q,i}) + 1$. To prove the lemma, we will carefully compare the behaviour of the sequences ψ_n and $\tilde{\psi}_n$. First, note that

$$\tilde{\psi}_0 \leq \psi_2 \leq \tilde{\psi}_4 \tag{4.12}$$

provided i is sufficiently large (because then $\sigma_n, \tilde{\sigma}_n$ are small compared to ε and δ). Note also that since ρ_n is decreasing, we have

$$\sigma_{n+2} > \tilde{\sigma}_n \tag{4.13}$$

for every $n \geq 0$. By (4.11),

$$\psi_{n+3} - \tilde{\psi}_{n+1} = (\psi_{n+2} - \tilde{\psi}_n)(1 + \varepsilon(\psi_{n+2} + \tilde{\psi}_n)) + \sigma_{n+2}^2 - \tilde{\sigma}_n^2$$

for $n \leq \min(N_{p,i} - 2, N_{q,i})$, where $\varepsilon(\psi_{n+2} + \tilde{\psi}_n) < \varepsilon\pi/2 < 1$. Hence, by induction, using (4.12) and (4.13), we obtain

$$\psi_{n+2} \geq \tilde{\psi}_n \tag{4.14}$$

for $n \in [0, \min(N_{p,i} - 2, N_{q,i}) + 1]$. In particular,

$$N_{p,i} < N_{q,i} + 2 \quad \text{or} \quad \psi_{N_{q,i}+2} > \tilde{\psi}_{N_{q,i}} \geq \delta,$$

which gives

$$N_{p,i} \leq N_{q,i} + 2. \tag{4.15}$$

The proof of the opposite estimate is more involved, so let us first present its sketch. We fix a number k such that (roughly speaking) $\psi_k \approx 1/i$. Then we show inductively $\tilde{\psi}_{n+2} \geq \psi_n - cn/i^3$ for $n \leq k$ and some constant $c > 0$ (see (4.18)). This gives $\tilde{\psi}_{k+2} \geq \psi_k - c'/i^2$ for some $c' > 0$ (see (4.19)). By the definition of k , we check that for sufficiently large constant $M > 0$ we have $\tilde{\psi}_{k+M} \geq \psi_k + c''M/i^2$ for some $c'' > 0$. With this starting condition, we inductively show $\tilde{\psi}_{n+M} \geq \psi_n + c''M/i^2$ for $n \in [k, N_{p,i}]$ (see (4.23)). This provides $\tilde{\psi}_{N_{p,i}+M} \geq \psi_{N_{p,i}} \geq \delta$, so $N_{q,i} \leq N_{p,i} + M$.

Now let us go into the details of the proof. By Lemmas 4.8 and 4.9, we have

$$n_{p,i}^- \asymp n_{q,i}^- \asymp i^2, \quad N_{p,i} \asymp N_{q,i} \asymp i, \tag{4.16}$$

so by Lemma 4.6,

$$\sigma_n \leq \frac{c_1}{i}, \quad \sigma_n^2 - \tilde{\sigma}_{n+2}^2 = (\sigma_n + \tilde{\sigma}_{n+2})(\sigma_n - \tilde{\sigma}_{n+2}) \leq \frac{c_1}{i^3} \tag{4.17}$$

for $n \in [0, N_{q,i} + 4]$ and a constant $c_1 > 0$. Let

$$k = \max \left\{ n \in [2, N_{q,i}] : \psi_{n+4} < \frac{b}{i} \right\}$$

for a small constant $b > 0$ (to be specified later). Note that $k \leq \min(N_{p,i} - 5, N_{q,i})$, so (4.11) holds for $n \in [2, k]$.

We will show by induction that

$$\psi_n - \tilde{\psi}_{n+2} \leq \frac{2c_1n}{i^3} \tag{4.18}$$

for every $n \in [2, k]$. For $n = 2$, (4.18) holds due to (4.12). Suppose it holds for some $n \in [2, k)$. By (4.11), we have

$$\psi_{n+1} - \tilde{\psi}_{n+3} = (\psi_n - \tilde{\psi}_{n+2})(1 + \varepsilon(\psi_n + \tilde{\psi}_{n+2})) + \sigma_n^2 - \tilde{\sigma}_{n+2}^2,$$

where by (4.14) and the definition of k , $\psi_n + \tilde{\psi}_{n+2} \leq \psi_n + \psi_{n+4} < 2\psi_{n+4} < 2b/i$, so using (4.16), (4.17) and the inductive assumption (4.18), we obtain

$$\psi_{n+1} - \tilde{\psi}_{n+3} \leq \frac{2c_1n}{i^3} \left(1 + \frac{2\varepsilon b}{i} \right) + \frac{c_1}{i^3} \leq \left(2n + \frac{4\varepsilon b N_{q,i}}{i} + 1 \right) \frac{c_1}{i^3} < \frac{(2n + c_2b + 1)c_1}{i^3}$$

for some constant $c_2 > 0$. Choosing the constant b in the definition of k sufficiently small, we can assume $c_2b < 1$, which gives

$$\psi_{n+1} - \tilde{\psi}_{n+3} \leq \frac{2c_1(n+1)}{i^3}.$$

This completes the inductive step and proves (4.18).

By (4.16) and (4.18),

$$\tilde{\psi}_{k+2} \geq \psi_k - \frac{c_3}{i^2} \tag{4.19}$$

for a constant $c_3 > 0$, while (by the definition of k),

$$\psi_{k+5} \geq \frac{b}{i} \tag{4.20}$$

and by (4.11),

$$\psi_{k+5} = \psi_k + \varepsilon(\psi_k^2 + \dots + \psi_{k+4}^2) + \sigma_k^2 + \dots + \sigma_{k+4}^2 < \psi_k + \frac{5(\varepsilon b^2 + c_1)}{i^2}. \tag{4.21}$$

by the definition of k , (4.11) and (4.17). Using (4.19), (4.20) and (4.21), we obtain

$$\tilde{\psi}_{k+2} \geq \frac{b}{i} - \frac{5(\varepsilon b^2 + c_1) + c_3}{i^2} \geq \frac{b}{2i} \tag{4.22}$$

for large i .

Take a large constant $M > 0$. We will show inductively

$$\tilde{\psi}_{n+M} - \psi_n \geq \frac{M\varepsilon b^2}{5i^2} \tag{4.23}$$

for $n \in [k, N_{p,i}]$. By (4.11), (4.19) and (4.22), we have

$$\begin{aligned} \tilde{\psi}_{k+M} &\geq \tilde{\psi}_{k+2} + \varepsilon(\tilde{\psi}_{k+2}^2 + \dots + \tilde{\psi}_{k+M}^2) \geq \tilde{\psi}_{k+2} + (M-2)\varepsilon\tilde{\psi}_{k+2}^2 \\ &\geq \tilde{\psi}_{k+2} + \frac{(M-2)\varepsilon b^2}{4i^2} \geq \psi_k - \frac{c_3}{i^2} + \frac{(M-2)\varepsilon b^2}{4i^2} \geq \psi_k + \frac{M\varepsilon b^2}{5i^2}, \end{aligned}$$

if M is chosen sufficiently large, so (4.23) holds for $n = k$. Suppose (4.23) holds for some $n \in [k, N_{p,i}]$. Now (4.15) implies that (4.11) is valid for n , so

$$\tilde{\psi}_{n+1+M} - \psi_{n+1} = (\tilde{\psi}_{n+M} - \psi_n)(1 + \varepsilon(\tilde{\psi}_{n+M} + \psi_n)) + \tilde{\sigma}_{n+M}^2 - \sigma_n^2,$$

where

$$\tilde{\psi}_{n+M} + \psi_n > \tilde{\psi}_{k+M} + \psi_k > \tilde{\psi}_{k+2}$$

for large i by (4.20) and (4.21) (which imply $\psi_k > 0$), while

$$\tilde{\sigma}_{n+M}^2 - \sigma_n^2 > -\frac{c_4}{i^3}$$

for a constant $c_4 > 0$ by (4.16) and Lemma 4.6 (with estimates analogous to the ones in (4.17)). Hence, using (4.22) we obtain

$$\tilde{\psi}_{M+n+1} - \psi_{n+1} \geq \frac{M\varepsilon b^2}{5i^2}(1 + \varepsilon\tilde{\psi}_{k+2}) - \frac{c_4}{i^3} \geq \frac{M\varepsilon b^2}{5i^2} \left(1 + \frac{\varepsilon b}{2i}\right) - \frac{c_4}{i^3} \geq \frac{M\varepsilon b^2}{5i^2},$$

provided M is chosen sufficiently large. This ends the inductive step and proves (4.23).

By (4.23),

$$\tilde{\psi}_{N_{p,i}+M} \geq \psi_{N_{p,i}} \geq \delta,$$

so

$$N_{q,i} \leq N_{p,i} + M.$$

This and (4.15) end the proof of the lemma. \square

Remark 4.11. Proving Lemmas 4.8–4.10, we have made the calculations for the initial point $(x, y) = (r_0 \cos \varphi_0, r_0 \sin \varphi_0)$ assuming $r_0 \in (0, 1)$. In fact, the case $r_0 > 1$ can be treated analogously. This can be seen by noting that Φ is symmetric with respect to r around the circle $r = 1$, while the only properties of R used in the proofs of the lemmas are the ones stated in Lemma 4.4. As the initial terms of the Taylor expansion of R near $r = 1$ are symmetric around 1, we see that an analogue of Lemma 4.6 holds in the case $r_0 > 1$ and the proof of Lemmas 4.8–4.10 can be repeated in that case. We conclude that Lemmas 4.8–4.10 hold for every initial point $(x, y) \in \mathbb{S}^2 \setminus (S \cup \{(0, 0), \infty\})$.

We summarize the results of this subsection in the following proposition.

Proposition 4.12. *For every $(x, y) \in \mathbb{S}^2 \setminus (S \cup \{(0, 0), \infty\})$ and every $\delta > 0$, if $N_{p,i}(x, y)$ (resp. $N_{q,i}(x, y)$) is the time spent by the trajectory of (x, y) under f during its i -th visit in the δ -neighbourhood U_p of p (resp. U_q of q), defined in (4.2), then*

$$N_{p,i}(x, y) \asymp N_{q,i}(x, y) \asymp i$$

and

$$|N_{p,i}(x, y) - N_{q,i}(x, y)| \leq M$$

for some constant $M > 0$, while the times spent by the trajectory between consecutive visits in $U_p \cup U_q$ are uniformly bounded.

This implies the following.

Corollary 4.13. *For every $(x, y) \in \mathbb{S}^2 \setminus (S \cup \{(0, 0), \infty\})$,*

$$\lim_{m \rightarrow \infty} \frac{1}{m} \sum_{n=0}^{m-1} \delta_{f^n(x,y)} = \frac{1}{2} \delta_p + \frac{1}{2} \delta_q$$

in the sense of weak-* convergence.

Proof. Fix $(x, y) \in \mathbb{S}^2 \setminus (S \cup \{(0, 0), \infty\})$ and $\delta > 0$. It is sufficient to prove that for the δ -neighbourhoods U_p and U_q , defined in (4.2), one has

$$\lim_{m \rightarrow \infty} \frac{1}{m} \sum_{n=0}^{m-1} \mathbb{1}_{U_p}(f^n(x, y)) = \lim_{m \rightarrow \infty} \frac{1}{m} \sum_{n=0}^{m-1} \mathbb{1}_{U_q}(f^n(x, y)) = \frac{1}{2}.$$

Fix $m \in \mathbb{N}$ and let $i = i(m)$ be the number of visits of (x, y) to U_p completed up to the time m , i.e. let i be the unique number such that

$$n_{p,i}^- \leq m < n_{p,i+1}^-.$$

Then by Proposition 4.12, there exist a constant $c > 0$ (independent of m) such that

$$\frac{i^2}{c} \leq \sum_{n=0}^{m-1} \mathbb{1}_{U_p}(f^n(x, y)) \leq ci^2, \quad \frac{i^2}{c} \leq \sum_{n=0}^{m-1} \mathbb{1}_{U_q}(f^n(x, y)) \leq ci^2,$$

and

$$\sum_{n=0}^{m-1} \mathbb{1}_{(U_p \cup U_q)^c}(f^n(x, y)) \leq ci.$$

This implies

$$\frac{2i^2}{c} \leq m \leq 3ci^2 \tag{4.24}$$

provided i is large enough (which holds if m is large enough). Therefore,

$$\lim_{m \rightarrow \infty} \frac{1}{m} \sum_{n=0}^{m-1} \mathbb{1}_{(U_p \cup U_q)^c}(f^n(x, y)) = 0$$

and hence

$$\lim_{m \rightarrow \infty} \left(\frac{1}{m} \sum_{n=0}^{m-1} \mathbb{1}_{U_p}(f^n(x, y)) + \frac{1}{m} \sum_{n=0}^{m-1} \mathbb{1}_{U_q}(f^n(x, y)) \right) = 1. \tag{4.25}$$

Proposition 4.12 together with (4.24) implies

$$\left| \frac{1}{m} \sum_{n=0}^{m-1} \mathbb{1}_{U_p}(f^n(x, y)) - \frac{1}{m} \sum_{n=0}^{m-1} \mathbb{1}_{U_q}(f^n(x, y)) \right| \leq \frac{C}{i}$$

for a constant $C > 0$ (independent of m), hence

$$\lim_{m \rightarrow \infty} \left| \frac{1}{m} \sum_{n=0}^{m-1} \mathbb{1}_{U_p}(f^n(x, y)) - \frac{1}{m} \sum_{n=0}^{m-1} \mathbb{1}_{U_q}(f^n(x, y)) \right| = 0. \tag{4.26}$$

Combining (4.25) with (4.26) finishes the proof (it is enough to notice that if a_n, b_n are sequences of real numbers with $\lim_{n \rightarrow \infty} (a_n + b_n) = 1$ and $\lim_{n \rightarrow \infty} |a_n - b_n| = 0$, then $\lim_{n \rightarrow \infty} a_n = \lim_{n \rightarrow \infty} b_n = \frac{1}{2}$). \square

4.3. Construction of the diffeomorphism $T : \mathbb{S}^2 \times \mathbb{S}^1 \rightarrow \mathbb{S}^2 \times \mathbb{S}^1$. Let

$$X = \mathbb{S}^2 \times \mathbb{S}^1,$$

where $\mathbb{S}^2 \simeq \mathbb{R}^2 \cup \{\infty\}$ and $\mathbb{S}^1 \simeq \mathbb{R}/\mathbb{Z}$. We can assume $X \subset \mathbb{R}^N$ for some $N \in \mathbb{N}$. Let

$$R_\alpha : \mathbb{S}^1 \rightarrow \mathbb{S}^1, \quad R_\alpha(t) = t + \alpha \pmod{1}, \quad \alpha \in \mathbb{R} \setminus \mathbb{Q}$$

be an irrational rotation. Recall that the normalized Lebesgue measure on \mathbb{S}^1 is the unique R_α -invariant Borel probability measure. Let

$$g : \mathbb{S}^1 \rightarrow \mathbb{S}^1, \quad g(t) = t + \frac{1}{100} \sin^2(\pi t) \pmod{1}.$$

Note that g is a C^∞ -diffeomorphism of \mathbb{S}^1 with 0 as the unique fixed point. Moreover, $\lim_{n \rightarrow \infty} g^n(t) = 0$ for every $t \in \mathbb{S}^1$. Therefore, δ_0 is the unique g -invariant Borel probability measure. Let $f : \mathbb{S}^2 \rightarrow \mathbb{S}^2$ be the diffeomorphism defined in Sect. 4.2, with the invariant unit circle $S \subset \mathbb{S}^2$ and the fixed points $p, q \in S$. Fix a small $\delta > 0$ and consider the δ -neighbourhoods $U_p, U_q \subset \mathbb{S}^2$ of p and q , respectively, defined in (4.2). Let

$$T : X \rightarrow X, \quad T(z, t) = (f(z), h_z(t)), \quad z \in \mathbb{S}^2, t \in \mathbb{S}^1,$$

where h_z are diffeomorphisms of \mathbb{S}^1 depending smoothly on $z \in \mathbb{S}^2$, such that $h_z = g$ for $z \in U_p$, $h_z = R_\alpha$ for $z \in U_q$, and for z outside $U_p \cup U_q$, h_z is defined in any manner which makes T a C^∞ -diffeomorphism of X .²

In view of Corollary 4.3, to conclude the proof of Theorem 1.11, it is sufficient to show the following.

Theorem 4.14. *The map T has an attractor*

$$\Lambda = S \times \mathbb{S}^1$$

with the basin $B(\Lambda) = (\mathbb{S}^2 \setminus \{(0, 0), \infty\}) \times \mathbb{S}^1$ and natural measure

$$\mu = \frac{1}{2} \delta_{p_0} + \frac{1}{2} \text{Leb}_{\mathbb{S}^1},$$

where $p_0 = (p, 0)$ and $\text{Leb}_{\mathbb{S}^1}$ is the Lebesgue measure on the circle $\{q\} \times \mathbb{S}^1$.

Before proving Theorem 4.14 we show the following lemma.

Lemma 4.15. *Let $T : X \rightarrow X$ be a continuous transformation of a compact metric space. Let ν_n , $n \geq 0$, be a sequence of Borel probability measures on X and let $\mathcal{A} \subset \mathbb{N} \cup \{0\}$ be a set of asymptotic density zero, i.e.*

$$\lim_{m \rightarrow \infty} \frac{1}{m} \#\{0 \leq n < m : n \in \mathcal{A}\} = 0.$$

Assume $\nu_{n+1} = T_* \nu_n$ for $n \notin \mathcal{A}$. Then any weak-* limit point of the sequence

$$\frac{1}{m} \sum_{n=0}^{m-1} \nu_n$$

is T -invariant.

Proof. Let ν be a weak-* limit of a sequence $\frac{1}{m_k} \sum_{n=0}^{m_k-1} \nu_n$ for some sequence $m_k \nearrow \infty$.

Then

$$T_* \nu - \nu = \lim_{k \rightarrow \infty} \frac{1}{m_k} \sum_{n=0}^{m_k-1} (T_* \nu_n - \nu_n) \tag{4.27}$$

and we will prove

$$\lim_{k \rightarrow \infty} \left\| \frac{1}{m_k} \sum_{n=0}^{m_k-1} (T_* \nu_n - \nu_n) \mathbb{1}_{\mathcal{A}}(n) \right\| = 0 \tag{4.28}$$

² This is possible since g is smoothly isotopic to identity by the family of maps $g_\varepsilon(t) = t + \varepsilon \sin^2(\pi t) \bmod 1$, $\varepsilon \in [0, \frac{1}{100}]$, while R_α is smoothly isotopic to identity by the family of maps $R_\varepsilon(t) = t + \varepsilon$, $\varepsilon \in [0, \alpha]$.

and

$$\lim_{k \rightarrow \infty} \left\| \frac{1}{m_k} \sum_{n=0}^{m_k-1} (T_* v_n - v_n) \mathbb{1}_{\mathcal{A}^c}(n) \right\| = \lim_{k \rightarrow \infty} \left\| \frac{1}{m_k} \sum_{n=0}^{m_k-1} (v_{n+1} - v_n) \mathbb{1}_{\mathcal{A}^c}(n) \right\| = 0, \tag{4.29}$$

where $\| \cdot \|$ stands for the total variation norm. Due to (4.27), this will imply $T_* v = v$. For (4.28), we have

$$\lim_{k \rightarrow \infty} \left\| \frac{1}{m_k} \sum_{n=0}^{m_k-1} (T_* v_n - v_n) \mathbb{1}_{\mathcal{A}}(n) \right\| \leq \lim_{k \rightarrow \infty} \frac{2}{m_k} \sum_{n=0}^{m_k-1} \mathbb{1}_{\mathcal{A}}(n) = 0,$$

as the asymptotic density of \mathcal{A} is zero and all v_n and $T_* v_n$ are probability measures. For (4.29), observe that the first equality follows by assumptions, and for a given $n \in \{0, \dots, m_k - 2\}$, if both n and $n + 1$ are in \mathcal{A}^c , then v_{n+1} cancels out in the sum $\sum_{n=0}^{m_k-1} (v_{n+1} - v_n) \mathbb{1}_{\mathcal{A}^c}(n)$ and otherwise it appears in the above sum at most once (possibly with a negative sign). The terms v_0 and v_{m_k} appear at most once. Therefore,

$$\begin{aligned} & \lim_{k \rightarrow \infty} \left\| \frac{1}{m_k} \sum_{n=0}^{m_k-1} (v_{n+1} - v_n) \mathbb{1}_{\mathcal{A}^c}(n) \right\| \\ & \leq \lim_{k \rightarrow \infty} \frac{1}{m_k} \left(\|v_{m_k}\| + \|v_0\| + \sum_{n=0}^{m_k-2} \|v_{n+1}\| (1 - \mathbb{1}_{\mathcal{A}^c}(n) \mathbb{1}_{\mathcal{A}^c}(n+1)) \right) \\ & = \lim_{k \rightarrow \infty} \frac{1}{m_k} \left(2 + \sum_{n=0}^{m_k-2} (1 - \mathbb{1}_{\mathcal{A}^c}(n) \mathbb{1}_{\mathcal{A}^c}(n+1)) \right) \\ & \leq \lim_{k \rightarrow \infty} \frac{1}{m_k} \left(2 + \sum_{n=0}^{m_k-2} (\mathbb{1}_{\mathcal{A}}(n) + \mathbb{1}_{\mathcal{A}}(n+1)) \right) = 0. \end{aligned}$$

□

Let us proceed now with the proof of Theorem 4.14.

Proof of Theorem 4.14. By the construction of f , the set Λ is a compact T -invariant set, and for every $(z, t) \in (\mathbb{S}^2 \setminus \{(0, 0), \infty\}) \times \mathbb{S}^1$, we have $\text{dist}(T^n(z, t), \Lambda)$ as $n \rightarrow \infty$. Hence, Λ is an attractor for T with the basin $B(\Lambda) = (\mathbb{S}^2 \setminus \{(0, 0), \infty\}) \times \mathbb{S}^1$. To prove that μ is a natural measure for T , we show that the sequence of measures

$$\mu_m = \frac{1}{m} \sum_{n=0}^{m-1} \delta_{T^n(z,t)}$$

converges to μ in the weak- $*$ topology for every $(z, t) \in (\mathbb{S}^2 \setminus (S \cup \{(0, 0), \infty\})) \times \mathbb{S}^1$. It is enough to prove that every limit point of the sequence μ_m is equal to μ . It follows from Corollary 4.13 that every such limit point must be of the form $v_1/2 + v_2/2$, where v_1 is a probability measure on the circle $\{p\} \times \mathbb{S}^1$ and v_2 is a probability measure on the circle $\{q\} \times \mathbb{S}^1$. Our goal is to show that $v_1 = \delta_{(p,0)}$ and $v_2 = \text{Leb}_{\mathbb{S}^1}$, where $\text{Leb}_{\mathbb{S}^1}$ is the Lebesgue measure on $\{q\} \times \mathbb{S}^1$.

Take $m_k \nearrow \infty$ such that $\lim_{k \rightarrow \infty} \mu_{m_k} = \nu_1/2 + \nu_2/2$. Let

$$\vartheta_{p,k} = \frac{1}{m_k} \sum_{n=0}^{m_k-1} \mathbb{1}_{U_p}(f^n(z)) \delta_{T^n(z,t)}, \quad \vartheta_{q,k} = \frac{1}{m_k} \sum_{n=0}^{m_k-1} \mathbb{1}_{U_q}(f^n(z)) \delta_{T^n(z,t)}$$

and

$$\vartheta_{O,k} = \frac{1}{m_k} \sum_{n=0}^{m_k-1} \mathbb{1}_{\mathbb{S}^2 \setminus (S \cup \{(0,0), \infty\}) \cup U_p \cup U_q}(f^n(z)) \delta_{T^n(z,t)}.$$

Clearly,

$$\mu_{m_k} = \vartheta_{p,k} + \vartheta_{q,k} + \vartheta_{O,k}.$$

By Corollary 4.13,

$$\lim_{k \rightarrow \infty} \vartheta_{p,k} = \frac{1}{2} \nu_1, \quad \lim_{k \rightarrow \infty} \vartheta_{q,k} = \frac{1}{2} \nu_2 \quad \text{and} \quad \lim_{k \rightarrow \infty} \vartheta_{O,k} = 0.$$

Let

$$\pi : X \rightarrow \mathbb{S}^1, \quad \pi(z, t) = t$$

be the projection. As $\text{supp } \nu_1 \subset \{p\} \times \mathbb{S}^1$ and $\text{supp } \nu_2 \subset \{q\} \times \mathbb{S}^1$ and g, R_α are uniquely ergodic with invariant measures δ_0 and $\text{Leb}_{\mathbb{S}^1}$, respectively, it is enough to show that the limits of projected measures $\pi_* \vartheta_{p,k}$ and $\pi_* \vartheta_{q,k}$ are, respectively, g and R_α -invariant.

We have

$$\pi_* \vartheta_{p,k} = \frac{1}{m_k} \sum_{n=0}^{m_k-1} \mathbb{1}_{U_p}(f^n(z)) \delta_{\pi(T^n(z,t))},$$

Let

$$M_k = \sum_{n=0}^{m_k-1} \mathbb{1}_{U_p}(f^n(z))$$

be the number of iterates $f^n(z)$ which are in U_p up to time m_k-1 and let $(z_0, t_0), (z_1, t_1), \dots$ be consecutive elements of the trajectory $\{T^n(z, t)\}_{n=0}^\infty$, such that $(z_j, t_j) \in U_p \times \mathbb{S}^1$. Then

$$\pi_* \vartheta_{p,k} = \frac{1}{m_k} \sum_{j=0}^{M_k-1} \delta_{t_j}.$$

Note that if $f(z_j) \in U_p$, then $t_{j+1} = g(t_j)$, so $\delta_{t_{j+1}} = g_* \delta_{t_j}$. Let $\mathcal{A} = \{j \in \mathbb{N} : f(z_j) \notin U_p\}$. By Proposition 4.12, the set \mathcal{A} has asymptotic density zero, as the time spent in U_p by the trajectory of z under f during its i -th visit grows linearly with i , while during each visit only the last iterate is such that $f(z_j) \notin U_p$. We can therefore apply

Lemma 4.15 to conclude that the sequence $\frac{1}{M_k} \sum_{j=0}^{M_k-1} \delta_{t_j}$ converges to a g -invariant probability measure, hence

$$\lim_{k \rightarrow \infty} \frac{1}{M_k} \sum_{j=0}^{M_k-1} \delta_{t_j} = \delta_0.$$

On the other hand, Corollary 4.13 implies $\lim_{k \rightarrow \infty} \frac{M_k}{m_k} = \frac{1}{2}$, so

$$\lim_{k \rightarrow \infty} \pi_* \vartheta_{p,k} = \frac{1}{2} \delta_0.$$

By the same arguments we show

$$\lim_{k \rightarrow \infty} \pi_* \vartheta_{q,k} = \frac{1}{2} \text{Leb}_{\mathbb{S}^1}.$$

Therefore, μ_m converges to μ in the weak- $*$ topology and μ is a natural measure for T . \square

Remark 4.16. To obtain a counterexample to the SSOY predictability conjecture in its original formulation, one can also perform a similar construction on a manifold with boundary $\mathbb{B} \times \mathbb{S}^1$, where \mathbb{B} is a closed 2-dimensional disc. Namely, it is enough to replace the diffeomorphism f of \mathbb{S}^2 constructed in Sect. 4.2 with a diffeomorphism of \mathbb{B} , which is a suitable modification if the ‘Bowen’s eye’ example described e.g. in [Cat14, Example 5.2.(B)], with properties similar to f .

Acknowledgements We are grateful to Edward Ott for bringing the paper [SSOY98] to our attention and to Balázs Bárány for informing us about the results of [Sim12]. We also thank Károly Simon for useful discussions. KB and AŚ were partially supported by the National Science Centre (Poland) grant 2019/33/N/ST1/01882. YG was partially supported by the National Science Centre (Poland) grant 2020/39/B/ST1/02329.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

[BGŚ20] Barański, K., Gutman, Y., Śpiewak, A.: A probabilistic Takens theorem. *Nonlinearity* **33**(9), 4940–4966 (2020)

[Cab00] Caballero, V.: On an embedding theorem. *Acta Math. Hungar.* **88**(4), 269–278 (2000)

[Cat14] Catsigeras, E.: On Ilyashenko’s statistical attractors. *Dyn. Syst.* **29**(1), 78–97 (2014)

[CG93] Carleson, L., Gamelin, T.W.: *Complex dynamics*. Tracts in Mathematics. Springer-Verlag, New York, Universitext (1993)

[Fal97] Falconer, K.: *Techniques in Fractal Geometry*. Wiley, Chichester (1997)

[Fal04] Falconer, K.: *Fractal Geometry: Mathematical Foundations and Applications*. Wiley, Chichester (2004)

- [FLR02] Fan, A.-H., Lau, K.-S., Rao, H.: Relationships between different dimensions of a measure. *Monatsh. Math.* **135**(3), 191–201 (2002)
- [Fol99] Folland, G.B.: *Real analysis. Pure and Applied Mathematics (New York)*, 2nd edn. Wiley, New York (1999)
- [FOY83] Farmer, J.D., Ott, E., Yorke, J.A.: The dimension of chaotic attractors. *Phys. D* **7**(1–3), 153–180 (1983)
- [FS87] Farmer, J.D., Sidorowich, J.J.: Predicting chaotic time series. *Phys. Rev. Lett.* **59**, 845–848 (1987)
- [GQS18] Gutman, Y., Qiao, Y., Szabó, G.: The embedding problem in topological dynamics and Takens’ theorem. *Nonlinearity* **31**(2), 597–620 (2018)
- [Gut16] Gutman, Y.: Taken’s embedding theorem with a continuous observable. In: *Ergodic Theory*, pp. 134–141. De Gruyter, Berlin (2016)
- [HBS15] Hamilton, F., Berry, T., Sauer, T.: Predicting chaotic time series with a partial model. *Phys. Rev. E* **92**, 010902 (2015)
- [HGLS05] Hsieh, C.-H., Glaser, S.M., Lucas, A.J., Sugihara, G.: Distinguishing random environmental fluctuations from ecological catastrophes for the North Pacific Ocean. *Nature* **435**(7040), 336–340 (2005)
- [HSY92] Hunt, B.R., Sauer, T., Yorke, J.A.: Prevalence: a translation-invariant “almost every” on infinite-dimensional spaces. *Bull. Amer. Math. Soc. (N.S.)* **27**(2), 217–238 (1992)
- [HW41] Hurewicz, W., Wallman, H.: *Dimension Theory*. Princeton Mathematical Series, v. 4. Princeton University Press, Princeton (1941)
- [Kec95] Kechris, A.S.: *Classical descriptive set theory*. Graduate Texts in Mathematics, vol. 156. Springer-Verlag, New York (1995)
- [KY90] Kostelich, E.J., Yorke, J.A.: Noise reduction: finding the simplest dynamical system consistent with the data. *Phys. D* **41**(2), 183–196 (1990)
- [Liu10] Liu, Z.: Chaotic time series analysis. *Math. Probl. Eng.*, pages Art. ID 720190, 31 (2010)
- [Mat95] Mattila, P.: *Geometry of sets and measures in Euclidean spaces*. Cambridge Studies in Advanced Mathematics, vol. 44. Cambridge University Press, Cambridge (1995)
- [Mil06] Milnor, J.: *Dynamics in one complex variable*, volume 160 of *Annals of Mathematics Studies*, 3rd edn. Princeton University Press, Princeton (2006)
- [MS04] McSharry, P.E., Smith, L.A.: Consistent nonlinear dynamics: identifying model inadequacy. *Physica D* **192**(1), 1–22 (2004)
- [NV20] Navarrete, R., Viswanath, D.: Prevalence of delay embeddings with a fixed observation function. *Phys. D* **414**, 132697, 15 (2020)
- [OL98] Ortega, G.J., Louis, E.: Smoothness implies determinism in time series: A measure based approach. *Phys. Rev. Lett.* **81**, 4345–4348 (1998)
- [Ott02] Ott, E.: *Chaos in Dynamical Systems*, 2nd edn. Cambridge University Press, Cambridge (2002)
- [OY08] Ott, W., Yorke, J.A.: When Lyapunov exponents fail to exist. *Phys. Rev. E* (3) **78**(5), 056203, 6 (2008)
- [PCFS80] Packard, N.H., Crutchfield, J.P., Farmer, J.D., Shaw, R.S.: Geometry from a time series. *Phys. Rev. Lett.* **45**, 712–716 (1980)
- [PdM82] Palis, J., Jr., de Melo, W.: *Geometric Theory of Dynamical Systems*. Springer-Verlag, New York-Berlin (1982)
- [QMAV99] Le Van, M., Quyen, J.M., Adam, C., Varela, F.J.: Nonlinear analyses of interictal EEG map the brain interdependences in human focal epilepsy. *Physica D* **127**(3), 250–266 (1999)
- [Rob05] Robinson, J.C.: A topological delay embedding theorem for infinite-dimensional dynamical systems. *Nonlinearity* **18**(5), 2135–2143 (2005)
- [Rob11] Robinson, J.C.: *Dimensions, Embeddings, and Attractors*. Cambridge Tracts in Mathematics, vol. 186. Cambridge University Press, Cambridge (2011)
- [Roh52] Rohlin, V.A.: On the fundamental ideas of measure theory. *Amer. Math. Soc. Translation* **1952**(71), 55 (1952)
- [SBDH97] Stark, J., Broomhead, D.S., Davies, M.E., Huke, J.P.: Takens embedding theorems for forced and stochastic systems. In: *Proceedings of the Second World Congress of Nonlinear Analysts, Part 8 (Athens, 1996)*, volume 30, pp. 5303–5314 (1997)
- [SBDH03] Stark, J., Broomhead, D.S., Davies, M.E., Huke, J.P.: Delay embeddings for forced systems. II. Stochastic forcing. *J. Nonlinear Sci.* **13**(6), 519–577 (2003)
- [SGM90] Sugihara, G., Grenfell, B., May, R.: Distinguishing error from chaos in ecological time-series. *Philos. Trans. R. Soc. B-Biol. Sci.* **330**(1257), 235–251 (1990)
- [Sim12] Simmons, D.: Conditional measures and conditional expectation; Rohlin’s disintegration theorem. *Discrete Contin. Dyn. Syst.* **32**(7), 2565–2582 (2012)
- [SM90] Sugihara, G., May, R.: Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series. *Nature* **344**(6268), 734–741 (1990)

- [SSOY98] Schroer, C.G., Sauer, T., Ott, E., Yorke, J.A.: Predicting chaos most of the time from embeddings with self-intersections. *Phys. Rev. Lett.* **80**, 1410–1413 (1998)
- [Sta99] Stark, J.: Delay embeddings for forced systems. I. Deterministic forcing. *J. Nonlinear Sci.* **9**(3), 255–332 (1999)
- [SYC91] Sauer, T.D., Yorke, J.A., Casdagli, M.: Embedology. *J. Statist. Phys.* **65**(3–4), 579–616 (1991)
- [Tak81] Takens, F.: Detecting strange attractors in turbulence. In: *Dynamical Systems and Turbulence*, Warwick 1980, volume 898 of *Lecture Notes in Math.*, pp. 366–381. Springer, Berlin-New York (1981)
- [Vos03] Voss, H.U.: Synchronization of reconstructed dynamical systems. *Chaos* **13**(1), 327–334 (2003)
- [Whi36] Whitney, H.: Differentiable manifolds. *Ann. Math. (2)* **37**(3), 645–680 (1936)
- [WV10] Yihong, W., Verdú, S.: Rényi information dimension: fundamental limits of almost lossless analog compression. *IEEE Trans. Inform. Theory* **56**(8), 3721–3748 (2010)
- [You02] Young, L.-S.: What are SRB measures, and which dynamical systems have them? *J. Statist. Phys.* **108**(5–6), 733–754 (2002)

Communicated by C.Liverani