

# On the timescale at which statistical stability breaks down<sup>☆</sup>

Neil Dobbs<sup>a,1</sup>, Alexey Korepanov<sup>b,\*,2</sup>

<sup>a</sup> University College Dublin, Ireland

<sup>b</sup> Exeter University, UK

Received 24 July 2018; received in revised form 11 June 2020; accepted 11 June 2020

Available online 18 June 2020

## Abstract

In dynamical systems, understanding statistical properties shared by most orbits and how these properties depend on the system are basic and important questions. Statistical properties may persist as one perturbs the system (*statistical stability* is said to hold), or may vary wildly. The latter case is our subject of interest, and we ask at what timescale does statistical stability break down. This is the time needed to observe, with a certain probability, a substantial difference in the statistical properties as described by (large but finite time) Birkhoff averages.

The quadratic (or logistic) family is a natural and fundamental example where statistical stability does not hold. We study this family. When the base parameter is of Misiurewicz type, we show, sharply, that if the parameter changes by  $t$ , it is necessary and sufficient to observe the system for a time at least of the order of  $|t|^{-1}$  to see the lack of statistical stability.

© 2020 L'Association Publications de l'Institut Henri Poincaré. Published by Elsevier B.V. All rights reserved. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

**Keywords:** Physical measures; Statistical stability; Quadratic maps; Logistic maps; Birkhoff averages; Finite timescale

## 1. Introduction

In this paper, we investigate the timescale at which statistical stability of dynamical systems breaks down. We carry out this study in the quadratic family, a standard test-bed for new directions in dynamics. The main theorems are stated in §2.

A real-world system can be represented by a phase space  $X$ , the set of all possible configurations of the system. Its evolution, with discrete time-steps, is described by a map  $f: X \rightarrow X$ . Suppose  $X$  is a Riemannian manifold and  $f$

<sup>☆</sup> The authors are very grateful to the referees for their thorough reading and helpful comments.

\* Corresponding author.

*E-mail addresses:* [neil.dobbs@ucd.ie](mailto:neil.dobbs@ucd.ie) (N. Dobbs), [a.korepanov@exeter.ac.uk](mailto:a.korepanov@exeter.ac.uk) (A. Korepanov).

<sup>1</sup> N.D. was supported by the ERC Bridges grant (ERC AdG 290843) while at the University of Geneva. This work sprang from the authors' encounter during the 2016 ESI programme "Mixing Flows and Averaging Methods" in Vienna.

<sup>2</sup> A.K. was funded by a European Advanced Grant *StochExtHomog* (ERC AdG 320977) at the University of Warwick and an Engineering and Physical Sciences Research Council grant EP/P034489/1 at the University of Exeter. A.K. is grateful to Mark Holland and Ian Melbourne for support and to University College Dublin for hospitality during his visit.

is continuous. If  $x, y \in X$  are nearby points, their orbits  $x, f(x), f^2(x), \dots$  and  $y, f(y), \dots$  remain close for a time. If the map is expanding, these orbits diverge in a time of the order of  $\log \text{dist}(x, y)^{-1}$  and may have very different properties. It is then natural to look at statistical properties of orbits, for example by studying Birkhoff averages

$$\bar{S}_n \varphi(x) = \frac{1}{n} \sum_{j=0}^{n-1} \varphi \circ f^j(x),$$

where  $\varphi: X \rightarrow \mathbb{R}$  is a continuous function (called an *observable*).

Perhaps surprisingly, in well-behaved systems, for a given  $\varphi$ , the Birkhoff averages may converge as  $n \rightarrow \infty$  to the same limit for almost every  $x$  with respect to the volume measure on  $X$ . Better still, there is a unique  $f$ -invariant probability measure  $\mu$  with the property that the limit is  $\int \varphi d\mu$  for every continuous  $\varphi$ .

### 1.1. Structural stability

Suppose we have a smooth one-parameter family of (discrete-time) maps  $f_t: X \rightarrow X$  for  $t$  in a neighbourhood of 0. The dynamics of nearby maps is relevant to the resilience to perturbation or if there is some uncertainty as to the governing parameters. If

$$\text{dist}(f_t(x), f_0(x)) \approx t,$$

as is reasonable, the orbits of  $x$  under  $f_0$  and  $f_t$  are expected to diverge in approximately  $\log |t|^{-1}$  time steps. Thus, comparing orbits of the same point under nearby maps does not lead very far. To deal with this, Andronov and Pontryagin [4] introduced the notion of *structural stability*, when for each nearby map there exists a global homeomorphism which maps orbits of the nearby map to orbits of the original. This concept works well for flows on compact surfaces [30,37] and more general Morse-Smale systems, for example.

Structural stability is a rather rigid property. A fundamental example where it fails is the family of quadratic (or *logistic*) maps

$$f_t: x \mapsto x^2 + (a + t),$$

where  $a + t$  lies in the parameter interval  $[-2, 1/4]$ . From Jakobson's Theorem [17], one deduces that the topological entropy of  $f_t$  is not locally constant at  $t = 0$  for any  $a$  in a positive-measure set of parameters. In particular, structural stability does not hold.

### 1.2. Statistical stability

Even without structural stability, statistical properties may appear to persist. Suppose that  $X$  is compact and let  $m$  denote the volume measure on  $X$ , normalized so that  $m(X) = 1$ . An  $f$ -invariant probability measure  $\mu$  on  $X$  is called *physical*, or *Sinai-Ruelle-Bowen (SRB)*, if there exists  $A \subset X$  with  $m(A) > 0$  so that for all continuous  $\varphi: X \rightarrow \mathbb{R}$  and  $x \in A$ ,

$$\lim_{n \rightarrow \infty} \bar{S}_n \varphi(x) = \int \varphi d\mu.$$

If  $m(A) = 1$ , we say that  $\mu$  is a *global physical measure*.

We say that the family  $f_t$  is *statistically stable*, if for every  $f_t$  there exists a global physical measure  $\mu_t$ , and for each continuous  $\varphi: X \rightarrow \mathbb{R}$ ,

$$\lim_{t \rightarrow 0} \int \varphi d\mu_t = \int \varphi d\mu_0.$$

Statistical stability has been studied by Keller [18], Dolgopyat [12], Alves and Viana [3], Alves, Carvalho and Freitas [2], Freitas and Todd [13] and others. The study of higher regularity properties was driven by Ruelle and Baladi, see [32,33,6] and references therein.

In the quadratic family, statistical stability holds at hyperbolic parameters (those corresponding to maps with periodic attractors). However, it does not hold everywhere, failing at most non-hyperbolic parameters [38,11], even near the so-called Misiurewicz parameters [11]. Moreover, there are quadratic maps [16] for which there is no physical measure to begin with.

**Remark 1.1.** One can obtain highly non-trivial positive results concerning statistical stability [40,13], and even Hölder continuity of the map  $t \mapsto \int \varphi d\mu_t$  [6], if the parameter range is restricted to a nowhere dense, but positive measure, set.

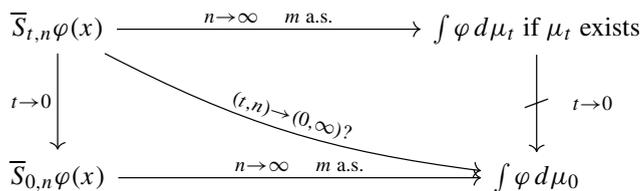
### 1.3. The breakdown of statistical stability

Introducing  $t$ -dependence to our Birkhoff averages, we set

$$\bar{S}_{t,n}\varphi = \frac{1}{n} \sum_{j=0}^{n-1} \varphi \circ f_t^j.$$

For each  $t, n$ , we view  $\bar{S}_{t,n}\varphi$  as random a variable on the probability space  $(X, m)$ . We suppose that  $f_0$  admits a global physical measure  $\mu_0$ , and we use  $\mu_t$  to refer to the global physical measures for  $f_t$ , whenever they exist.

Consider the following diagram.



Following the lower-left path,

$$\lim_{n \rightarrow \infty} \lim_{t \rightarrow 0} \bar{S}_{t,n}\varphi(x) = \int \varphi d\mu_0 \quad m\text{-almost surely.}$$

Switch the order of limits and this will no longer hold. The measures  $\mu_t$  need not exist, and even restricting to parameters for which they do, the integrals  $\int \varphi d\mu_t$  need not vary continuously.

Now consider the diagonal arrow. Let  $n(t)$  be an integer-valued function of  $t$  with  $n(t) \rightarrow \infty$  as  $t \rightarrow 0$ . Intuitively, if  $n(t) \ll -\log |t|$ , then orbits of a point  $x$  under  $f_0$  and under  $f_t$  do not have time to meaningfully diverge, so  $\bar{S}_{t,n(t)}\varphi \approx \bar{S}_{0,n(t)}\varphi$  and

$$\lim_{t \rightarrow 0} \bar{S}_{t,n(t)}\varphi = \int \varphi d\mu_0 \quad m \text{ almost-surely.} \tag{1.1}$$

As a corollary,

$$\bar{S}_{t,n(t)}\varphi \rightarrow \int \varphi d\mu_0 \quad \text{in probability (w.r.t. } m), \text{ as } t \rightarrow 0. \tag{1.2}$$

The almost sure convergence (1.1) is a rather rigid concept, it is expected to break down once  $n(t) \gg -\log |t|$ , see [21, Section 7].

In this paper, we examine how fast  $n(t)$  can grow without destroying the convergence in probability (1.2). Or, given the size of a small perturbation, we determine the minimum amount of observation time needed to discover instability in the statistical behaviour. Similarly, if we have some uncertainty in the parameter governing the system, the predicted statistical behaviour is valid up until some timescale.

For the quadratic family, if the base parameter is of Misiurewicz type, the statistical stability continues to hold as long as  $n(t)$  grows more slowly than  $t^{-1}$ , see Theorem 2.6. This result is sharp: if  $n(t)$  grows as fast as  $t^{-1}$ , continuity is lost, see Theorem 2.7. We say that, in this context,

*statistical stability breaks down at the timescale  $\frac{1}{t}$ .*

#### 1.4. Fast-slow systems

An initial stimulus for our work was the study of *fast-slow* systems of the form:

$$\begin{cases} s_{\varepsilon,n+1} = s_{\varepsilon,n} + \varepsilon\varphi(s_{\varepsilon,n}, x_{\varepsilon,n}), & s_{\varepsilon,0} = 0 \\ x_{\varepsilon,n+1} = f_{\varepsilon}(x_{\varepsilon,n}), & x_{\varepsilon,0} \sim m \end{cases} \quad (1.3)$$

with  $\varepsilon \in [0, \varepsilon_0]$ . When the maps  $f_{\varepsilon}$  are nonuniformly expanding, under rather general assumptions it is proved [21] that as  $\varepsilon \rightarrow 0$ , the random process  $s_{\varepsilon, \lfloor \varepsilon^{-1}t \rfloor}$ ,  $t \in [0, 1]$ , converges in distribution to the solution of the ordinary differential equation  $\dot{s} = \int \varphi(s, x) d\mu_0(x)$ ,  $s(0) = 0$ , where  $\mu_0$  is the physical measure for  $f_0$ .

In the case of logistic maps, to satisfy the assumption that the maps  $f_{\varepsilon}$  are nonuniformly expanding, the range of  $\varepsilon$  has to be restricted to a nowhere dense subset of  $[0, \varepsilon_0]$ . It is an interesting question whether the restriction on parameters can be removed. The authors of [21] were asked this question by various people, including D. Dolgopyat and the anonymous referee of [21].

To simplify the model, we suppose that  $\varphi$  does not depend on  $s$ , i.e.  $\varphi(s, x) = \varphi(x)$ . Then

$$s_{\varepsilon, \lfloor \varepsilon^{-1}t \rfloor} = \varepsilon \sum_{j=0}^{\lfloor \varepsilon^{-1}t \rfloor - 1} \varphi \circ f_{\varepsilon}^j.$$

Our theorems respond to the above question, showing that convergence breaks down without a restriction on the parameter range but, surprisingly, for all shorter (and less natural) timescales, one does have convergence.

#### 1.5. Stochastic stability

In this paper we perturb a dynamical system by considering another one close to the original. Such perturbations are called *deterministic*. Another type is *stochastic*, where at each step a small perturbation is chosen randomly.

Suppose the base map has a physical measure  $\mu_0$ . If the statistics of stochastically perturbed systems can be described by measures  $\mu_{\varepsilon}$ , where  $\varepsilon$  reflects the average strength of the perturbation, and if  $\mu_{\varepsilon} \rightarrow \mu_0$  as  $\varepsilon \rightarrow 0$ , then the base map is *stochastically stable*. The question of stochastic stability has been treated successfully in [1,5,7,9,18,24,35,36] among others.

In sharp contrast with statistical stability, almost every quadratic map is stochastically stable [7,9,35,25].

#### 1.6. Statistical detection of the lack of linear response

While preparing this manuscript for submission, we became aware of an interesting article [15] by Gottwald et al. which examines, via numerical experiments, the possibility of detecting statistically the lack of linear response in the quadratic family with a ‘global observable’.

Linear response is a stronger property than statistical stability. Even so, it is found in [15] that detection of absence of linear response requires a well-designed statistical test and observations on long timescales (such as  $10^6$  iterations).

The results of [15] suggest that one needs a timescale of order at least  $t^{-0.91}$  to detect the lack of linear response under perturbations of size  $t$  with a global observable. The timescale can be reduced by crafting special observables. These observations are compatible with our results concerning statistical stability. There is a more in-depth discussion in [15] about the implications for mathematical modelling.

#### 1.7. Organisation

The paper is organized as follows. In §2 we give formal definitions and statements of our main results. In §3 we assemble various results about the maps  $f_t$  close to the base map  $f_0$ . In §4 we study topological and metric properties of first return maps to carefully chosen small neighbourhoods of the critical point.

In §5 we prove the lack of statistical stability on the timescale  $n(t) = t^{-1}$ . We find parameters  $t_n$  with the critical point a super-attracting periodic point with period as short as possible. The size of the immediate basin of attraction of the critical point happens to be of the order of  $t_n$ . For any  $C > 0$ , we show that a definite proportion of points fall into the basin in fewer than  $Ct_n^{-1}$  iterates, which is enough to obliterate statistical stability.

In §6 we prove statistical stability on shorter timescales. There is a natural argument which works for timescales up to  $o(t^{-1/2})$  (see Remark 6.1), but this is not optimal. To reach the optimal  $o(t^{-1})$ , we intricately construct an induced map. We use it to approximate each  $f_t$  with a non-uniformly expanding map for which martingale approximations give strong control of statistical properties.

## 2. Statements

We shall often write  $Df$  for the derivative  $f'$  of a map  $f$ .

**Definition 2.1.** We say that a continuous map  $f : I \rightarrow I$ , defined on a compact interval  $I$ , is *unimodal* if  $f$  has exactly one turning point  $c$ . We say  $f$  is a *smooth unimodal map* if, moreover,  $f$  is continuously differentiable and  $c$  is the unique (*critical*) point satisfying  $f'(c) = 0$ . The critical point and the map are *non-degenerate* if  $f''(c) \neq 0$ .

**Definition 2.2.** A map  $f : I \rightarrow I$  is *S-unimodal* if it is a  $C^2$  smooth unimodal map with critical point  $c$ ,  $|f'|^{-1/2}$  is convex on each component of  $I \setminus \{c\}$ ,  $f(\partial I) \subset \partial I$  and  $|f'| > 1$  on  $\partial I$ .

The convexity condition is equivalent ([29], [10, p. 266]), for  $C^3$  maps, to having non-positive Schwarzian derivative, while strict convexity corresponds to negative Schwarzian derivative. Quadratic maps have negative Schwarzian derivative. A forward-invariant compact set  $X$  for  $f$  is *hyperbolic repelling* if there exists  $k \geq 1$  with  $|Df^k| \geq 2$  on  $X$ . The *post-critical orbit* is the set  $\{f^n(f(c))\}_{n \geq 0}$ .

**Definition 2.3.** A smooth unimodal map is called *Misiurewicz* if the closure of its post-critical orbit is a hyperbolic repelling set.

Misiurewicz maps have strong expansion properties which outweigh any contraction caused by passage close to the critical point. By Singer’s Theorem [10, Theorem III.1.6], all periodic points of an S-unimodal Misiurewicz map are hyperbolic repelling. We shall recall further properties anon.

Throughout the paper we fix  $I = [-1, 1]$ , and all our unimodal maps have the critical point at 0.

**Definition 2.4.** A *Misiurewicz-rooted unimodal family* is a family  $\{f_t\}_{t \in [0, \varepsilon]}$ ,  $\varepsilon > 0$ , of non-degenerate S-unimodal maps on  $I$  with the critical point 0. We require that  $f_0$  is a Misiurewicz map and  $f_t(x)$  is  $C^2$  as a function of  $(x, t)$ .

**Definition 2.5.** We say that a Misiurewicz-rooted unimodal family  $\{f_t\}$  is *transversal* if

$$\sum_{j=0}^{\infty} \frac{\partial_t f_t(f_0^j(0))|_{t=0}}{(f_0^j)'(f_0(0))} \neq 0.$$

Suppose that  $\{f_t\}$  is a Misiurewicz-rooted unimodal family and let  $\mu_0$  be the unique  $f_0$ -invariant absolutely continuous probability measure [28]. Let  $\varphi : I \rightarrow \mathbb{R}$  be a continuous observable and define

$$\bar{S}_{t,n}\varphi := \frac{1}{n} \sum_{j=0}^{n-1} \varphi \circ f_t^j.$$

Let  $\bar{\varphi} = \int \varphi d\mu_0$ . Let  $m$  denote the normalized to probability Lebesgue measure on  $I$ .

**Theorem 2.6 (Persistence of statistical stability).** For any function  $n : \mathbb{R}^+ \rightarrow \mathbb{Z}^+$  such that  $\lim_{t \rightarrow 0^+} n(t) = \infty$  and  $\lim_{t \rightarrow 0^+} tn(t) = 0$ ,

$$\lim_{t \rightarrow 0^+} \int_I |\bar{S}_{t,n(t)}\varphi - \bar{\varphi}| dm = 0.$$

**Theorem 2.7** (Breakdown of statistical stability). *Let  $a > 0$ . If the family  $\{f_t\}$  is transversal, then there exists a continuous observable  $\varphi$  for which*

$$\limsup_{t \rightarrow 0^+} \int_I \bar{S}_{t, \lfloor a/t \rfloor} \varphi \, dm \neq \bar{\varphi}.$$

**Remark 2.8.** One could ask whether being  $S$ -unimodal is necessary or whether just assuming  $C^2$  would suffice to prove these results. We principally use the  $S$ -unimodal convexity condition to simplify control of distortion. Mañé’s Hyperbolicity Theorem [26] for  $C^2$  maps gives expansion and distortion control for the dynamics outside a neighbourhood of the critical point. So another path exists, but it would take more work and we wished to avoid further complicating an already technical paper. Similarly, one could ask what happens for other, non-Misiurewicz, base parameters, for example Collet-Eckmann ones.

**Example 2.9.** Let  $g_t(x) = x^2 + t_0 + t$  be a parametrisation of the quadratic family, with  $t_0$  a Misiurewicz parameter in  $[-2, 1/4)$  and  $t \in [0, 1/4 - t_0)$ . Noting that  $\partial_t g_t(x) \equiv 1$ , transversality has been shown by Levin [23] (under more general summability conditions). This family does not leave  $[-1, 1]$  invariant, so it is not (quite) a Misiurewicz-rooted unimodal family. However, it can be transformed by a smooth family of affine transformations into a transversal Misiurewicz-rooted unimodal family. Hence our main theorems apply to the family  $g_t$ .

### 3. Preliminaries

We shall use the notation  $A(\cdot) = O(B(\cdot))$  and  $A(\cdot) \lesssim B(\cdot)$  interchangeably, meaning that there exists a constant  $C > 0$  such that  $A(\cdot) \leq CB(\cdot)$  for all sufficiently large (or small) values of the argument. If both  $A(\cdot) \lesssim B(\cdot)$  and  $B(\cdot) \lesssim A(\cdot)$ , we write  $A(\cdot) \simeq B(\cdot)$ .

**Definition 3.1.** Let  $W, V$  be open intervals. Suppose that  $g : W \rightarrow V$  is a  $C^2$  surjective diffeomorphism with  $|Dg|^{-1/2}$  convex. Suppose that  $g$  can be extended to a  $C^2$  surjective diffeomorphism  $g : \hat{W} \rightarrow \hat{V}$  with  $|Dg|^{-1/2}$  convex, where  $\hat{W}, \hat{V}$  are intervals and  $\hat{W}$  compactly contains  $W$ .

In this setup we say that  $g$  is  $\hat{W}$ -*extensible*. When both connected components of  $\hat{V} \setminus V$  have length at least  $\delta|V|$  for some  $\delta > 0$ , we say that  $g$  is  $\delta$ -*extensible*.

**Lemma 3.2** (Koebe Principle [10, Theorem IV.1.2], [28]). *Suppose that  $g : W \rightarrow V$  is a  $C^2$  surjective diffeomorphism with  $|Dg|^{-1/2}$  convex, and that  $g$  is  $\delta$ -extensible. Then we have the distortion bound*

$$\sup_{x, y \in W} \frac{Dg(x)}{Dg(y)} \leq \frac{(1 + \delta)^2}{\delta^2}.$$

In addition, there exists a constant  $C$  depending only on  $\delta$ , such that for all  $x, y \in W$ ,

$$|\log |Dg(x)| - \log |Dg(y)|| \leq \frac{C}{|W|} |x - y|.$$

Let us fix a constant  $\Delta > 1$  for which  $\Delta$ -extensible maps have distortion bounded by 2.

**Lemma 3.3.** *Suppose that  $g, W, \hat{W}, V, \hat{V}$  are as in Definition 3.1 and that, additionally, each component of  $\hat{V} \setminus V$  has length at least  $10(1 + \Delta)|V|$  and  $|V| > |\hat{W}|$ . Then  $|Dg| > 5$  on  $W$ .*

**Proof.** There is an interval  $V' \supset V$  with  $|V'| = 10|V|$ . Let  $W' = g^{-1}(V')$ . Each component of  $\hat{V} \setminus V'$  has length at least  $10\Delta|V| = \Delta|V'|$ , so  $g : W' \rightarrow V'$  is  $\Delta$ -extensible. By Lemma 3.3, the distortion of  $g$  is bounded by 2 on  $W'$ . The result then follows from the estimate  $|V'| = 10|V| > 10|W'|$ .  $\square$

Suppose that  $f : I \rightarrow I$  is a continuous map with  $f(\partial I) \subset \partial I$ .

**Definition 3.4.** We say that an interval  $A \subset I$  is a *pullback* of an interval  $U \subset I$  (under  $f$ ), if  $A$  is a connected component of  $f^{-n}(U)$  for some  $n \geq 0$ .

**Definition 3.5.** An open interval  $U$  is called *regularly returning* if  $f^n(\partial U) \cap U = \emptyset$  for all  $n \geq 0$ .

This property is widely used [14,27,31] to simplify the study of induced maps thanks to the following elementary property.

**Lemma 3.6.** *If  $U$  is regularly returning, then pullbacks of  $U$  are either nested or disjoint, that is, if  $A, B$  are pullbacks of  $U$  and if  $A \cap B \neq \emptyset$ , then either  $A \subset B$  or  $B \subset A$ .*

We shall use *induced maps* of the form  $F(x) = f^{\tau(x)}(x)$  in much of the paper, where  $\tau$  is an *inducing time*, defined on a disjoint union of open intervals, called *branches*, where  $\tau$  is constant. A branch is *full* if its image equals the range of the induced map.

*First entry* maps and *first return* maps to a regularly returning interval  $U$  are primary examples of induced maps. The first entry time is

$$e(x) = \inf\{k \geq 0 : f^k(x) \in U\},$$

while the first return time is

$$r(x) = \inf\{k \geq 1 : f^k(x) \in U\} = 1 + e(f(x)).$$

The first entry map  $x \mapsto f^{e(x)}(x)$  and the first return map  $x \mapsto f^{r(x)}(x)$  are defined on the sets  $\{x \in I : e(x) < \infty\}$  and  $\{x \in I : r(x) < \infty\}$  respectively.

Since  $U$  is regularly returning, it follows from Lemma 3.6 that if  $W$  is a branch of the first entry or the first return map with the corresponding inducing time  $n_W$ , then  $f^{n_W}(\partial W) \subset \partial U$ .

Henceforth, suppose that  $\{f_t\}$  is a Misiurewicz-rooted unimodal family. As a Misiurewicz map,  $f_0$  enjoys strong expansion properties:

**Lemma 3.7** ([10, Theorem III.6.3]). *Given any sufficiently small neighbourhood  $U$  of 0, there exist  $C \in (0, 1)$  and  $\lambda > 1$  such that for each  $x \in I$*

- if  $f_0^j(x) \notin U$  for  $0 \leq j \leq k - 1$ , then

$$|Df_0^k(x)| \geq C\lambda^k;$$

- if  $f_0^k(x) \in U$ , then

$$|Df_0^k(x)| \geq C\lambda^k.$$

The maps  $f_t$  for  $t \neq 0$  are not necessarily Misiurewicz, and Lemma 3.7 does not apply. Still, for  $t$  small enough, a similar statement holds:

**Lemma 3.8** ([10, Theorem III.6.4]). *There exists  $C \in (0, 1)$  and  $\lambda > 1$  such that, given any sufficiently small neighbourhood  $U$  of 0, the following holds for all sufficiently small  $t$ .*

- If  $f_t^j(x) \notin U$  for  $0 \leq j \leq k - 1$ , then

$$|Df_t^k(x)| \geq C\lambda^k \inf_{0 \leq j < k} |Df_t(f_t^j(x))|.$$

- If  $f_t^j(x) \notin U$  for  $0 \leq j \leq k - 1$  and  $f_t^k(x) \in U$ , then

$$|Df_t^k(x)| \geq C\lambda^k.$$

Expansion entails a uniform distortion bound.

**Lemma 3.9.** *Let  $U$  be a neighbourhood of 0. There is a constant  $C > 1$  such that, for all  $t$  small enough, the following holds. If  $W$  is an open interval such that  $f_t^k(W) \cap U = \emptyset$  for  $0 \leq k < n$  and  $x, y \in W$ , then*

$$|\log |Df_t^n(x)| - \log |Df_t^n(y)|| \leq (\log C) |f_t^n(x) - f_t^n(y)|.$$

**Proof.** By Lemma 3.8, there is a constant  $C_0 > 0$  (independent of  $t, W, n$ ) such that, for all  $x, y \in W$ ,

$$\sum_{k=0}^{n-1} |f_t^k(x) - f_t^k(y)| \leq C_0 |f_t^n(x) - f_t^n(y)|.$$

As  $f_0$  has a unique critical point at 0 and  $f_t(x)$  is a  $C^2$  function of  $(x, t)$ ,  $Df_t(x)$  is bounded away from 0 on  $I \setminus U$  and  $D^2 f_t$  is bounded. Consequently, there exists a constant  $C_1 > 0$ , depending on  $U$  but not on  $t$ , such that  $|D(\log |Df_t|)| \leq C_1$  on  $I \setminus U$ . For  $x, y \in W$ , we deduce

$$\begin{aligned} |\log |Df_t^n(y)| - \log |Df_t^n(x)|| &= \left| \sum_{k=0}^{n-1} \int_{f_t^k(x)}^{f_t^k(y)} D(\log |Df_t|)(z) dz \right| \\ &\leq C_1 \sum_{k=0}^{n-1} |f_t^k(x) - f_t^k(y)| \\ &\leq C_0 C_1 |f_t^n(x) - f_t^n(y)|. \quad \square \end{aligned}$$

The map  $f_0$ , being Misiurewicz, has an induced map with good properties.

**Lemma 3.10** ([10, Proof of Lemma V.3.2]). *For the map  $f_0$ , there is an arbitrarily small regularly-returning open interval  $J$  containing 0, disjoint from the post-critical orbit, for which  $f(\partial J)$  is a (single) periodic point. Each branch of the first return map is mapped diffeomorphically onto  $J$ . The complement in  $J$  of the domain of the first return map has zero Lebesgue measure. There is a uniform distortion bound for all iterates of the first return map.*

Let  $\theta_0 > 0$  be small enough that for any neighbourhood  $U$  of 0 contained in  $(-\theta_0, \theta_0)$ , the conclusions of Lemma 3.7 and Lemma 3.8 hold. We further require that  $\theta_0 < (10(1 + \Delta))^{-1}$ , the latter constant as in Lemma 3.3.

**Lemma 3.11.** *Let  $J \subset (-\theta_0, \theta_0)$  be an interval given by Lemma 3.10. Periodic points of  $f_0$  are dense in  $J$ . Preimages of any point in  $J$  are dense in  $J$  and hence in  $I$ .*

**Proof.** Let  $\phi: J \rightarrow J$  be the first return map to  $J$  under the iterations of  $f_0$ . The union of branches of  $\phi^n$  has full Lebesgue measure in  $J$  for each  $n$ . Because of the uniform distortion and expansion bounds given by Lemmas 3.10 and 3.7, the maximal diameter of a branch of  $\phi^n$  tends to 0 as  $n \rightarrow \infty$ .

Each branch  $A$  of  $\phi^n$  is mapped by  $\phi^n$  diffeomorphically onto  $J$ . Assuming that  $\partial A \cap \partial J = \emptyset$ , there is a point  $x \in A$  such that  $\phi^n(x) = x$ . Thus all but at most two branches of  $\phi^n$  contain a periodic point for  $f$ . It follows that periodic points are dense in  $J$ .

For  $x \in J$ , each branch of  $\phi^n$  contains a preimage of  $x$ , so the preimages are dense in  $J$ . Further, intervals  $A \subset I \setminus J$  such that  $f^n: A \rightarrow J$  is a diffeomorphism are dense in  $I$  by expansion outside  $J$  (see Lemma 3.7), so preimages of  $x$  are dense in  $I$ .  $\square$

Let  $\Lambda$  be a closed  $f_0$ -forward-invariant subset of  $I$  such that  $0 \notin \Lambda$ . We introduce the continuation of points in  $\Lambda$  (see [34, Lemma 3.1]).

**Lemma 3.12.** *There exist an integer  $N \geq 1$  and numbers  $\rho, t_0, C > 0$  such that the following holds for all  $t \in [0, t_0]$ . Given  $x \in \Lambda$ , there is a unique point  $x_t$  which satisfies  $|x_t - x| \leq Ct$  and  $|f_t^{Nj}(x_t) - f_0^{Nj}(x)| < \rho$  for all  $j \geq 0$ . The map  $t \mapsto x_t$  is continuous.*

**Proof.** By Lemma 3.8, for some  $N \geq 1$  and  $\rho > 0$ ,  $g_t := f_t^N$  satisfies  $|Dg_t| > 2$  on a  $\rho$ -neighbourhood  $B(\Lambda, \rho)$  of  $\Lambda$  for all  $t \in [0, t_1]$ , for some  $t_1 > 0$ . Recall that  $f_t(x)$  is a  $C^2$  function of  $(x, t)$ . Choose  $C > 1$  such that  $|g_t(x) - g_0(x)| \leq Ct$  for all  $x$  and all  $t \in [0, t_1]$ .

We now apply the Implicit Function Theorem. There exists  $t_0 \in (0, \min(t_1, \frac{\rho}{C}))$  such that, for each  $t \in [0, t_0]$  and  $x \in \Lambda$ , there is a unique  $y$  in the same connected component of  $B(\Lambda, \rho)$  as  $x$  which satisfies  $g_0(x) = g_t(y)$ . Moreover,

$$|x - y| \leq \frac{1}{2} |g_t(x) - g_t(y)| = \frac{1}{2} |g_t(x) - g_0(x)| \leq \frac{Ct}{2} < \frac{\rho}{2}.$$

Fix such  $x, t$  and  $y$ . If  $z \in B(g_0(x), Ct)$ , there is a unique  $y'$  in the same connected component as  $x$  of  $B(\Lambda, \rho)$  with  $z = g_t(y')$  and  $y'$  satisfies

$$|y' - x| \leq |y' - y| + |y - x| \leq \frac{1}{2} |g_t(y') - g_t(y)| + Ct/2 \leq Ct.$$

Inductively, for  $n \geq 0$  we obtain points  $y_n = y_n(t)$  such that  $g_t^n(y_n) = g_0^n(x)$  and  $|g_t^j(y_n) - g_0^j(x)| \leq Ct$  for  $j = 0, \dots, n$  and  $y_{n+1} \subset B(y_n, 2^{-n})$ .

In particular,  $(y_n)_n$  is a Cauchy sequence whose limit we denote by  $x_t$ . The point  $x_t$  satisfies  $|x_t - x| \leq Ct$  and  $|g_t^j(x_t) - g_0^j(x)| < \rho$  for all  $j \geq 0$ . Continuous dependence of  $x_t$  on  $t$  follows from continuous dependence of  $y_n(t)$  on  $t$ .  $\square$

In particular, for each  $x \in \Lambda$ , we obtain a map  $t \mapsto x_t$  with the same Lipschitz constant  $C$ . Combining them generates a map  $t \mapsto \Lambda_t$ . Note that if  $x$  is preperiodic for  $f_0$ , then from uniqueness it follows that  $x_t$  is preperiodic for  $f_t$ .

**Definition 3.13 (Continuation).** The map  $t \mapsto x_t$  as above (or the point  $x_t$ ) is called the *continuation* of  $x = x_0$ .  $\Lambda_t$  is called the *continuation* of  $\Lambda = \Lambda_0$ .

**Lemma 3.14.** Let  $\theta \in (0, \theta_0)$ . For sufficiently small  $t$ , there exist open intervals  $U_0, U_1$ , such that

- (a)  $0 \in U_1 \subset U_0 \subset (-\theta, \theta)$ ;
- (b) for each  $j$ , the boundary  $\partial U_j$  varies continuously with  $t$ , and  $f_t(\partial U_j)$  is a single point, preperiodic with respect to  $f_t$ ;
- (c)  $f_t^k(\partial U_j) \notin U_0$  for all  $k \geq 1$  and  $j = 0, 1$ ;
- (d)  $|U_1| \leq \theta \text{dist}(U_1, \partial U_0)$ .

**Proof.** Suppose first that  $t = 0$ . Let  $J \subset (-\theta/2, \theta/2)$  be given by Lemma 3.11, and set  $U_0 = J$ . Recall that  $f_0(\partial U_0)$  is a single periodic point whose orbit under  $f_0$  is disjoint from  $U_0$ .

Let  $F: U_0 \rightarrow U_0$  denote the first return map to  $U_0$  under  $f_0$ . Branches of  $F$  accumulate on 0, since 0 never returns, and boundary points of branches get mapped by the corresponding iterate of  $f_0$  to  $\partial U_0$ . Hence there are preperiodic points, arbitrarily close to 0, which never return to  $U_0$ . Choose one,  $p < 0$ , such that  $p$  and its symmetric point  $p_*$  (in the sense  $f_0(p) = f_0(p_*)$ ) lie in  $U_0$  and such that

$$|p_* - p| < \theta \text{dist}((p, p_*), \partial U_0)/2,$$

and set  $U_1 = (p, p_*)$ .

The boundaries of  $U_j$ ,  $j = 0, 1$ , consist of preperiodic points whose forward orbits do not include 0, hence they admit continuations, giving the sets  $U_j$  with the required properties for small enough  $t$ .  $\square$

**Lemma 3.15.** Let  $U_j$  denote the intervals from Lemma 3.14. Let

$$E_n = \{x \in I : f_t^k(x) \notin U_1 \text{ for all } k = 0, 1, 2, \dots, n\},$$

$$R_n = \{x \in I : f_t^k(x) \notin U_1 \text{ for all } k = 1, 2, \dots, n\}.$$

For  $t$  small enough, there are constants  $\alpha, C > 0$  such that

$$m(E_n) < C e^{-\alpha n} \quad \text{and} \quad m(R_n) < C e^{-\alpha n}$$

for all  $n \geq 0$ .

**Proof.** Choose a neighbourhood of 0 contained in  $U_1$  for all small  $t$  and obtain a distortion bound  $C' > 1$  from Lemma 3.9. Let us drop the dependence on  $t$  from notation, where appropriate.

Note that  $E_n$  is a finite union of closed intervals and  $E_{n+1} \subset E_n$ . Let  $A$  be a connected component of  $E_n$ . Then  $f^n$  is monotone on  $A$  and the boundary points of the interval  $f^n(A)$  are distinct elements of the preperiodic forward orbit of  $\partial U_1$ . Therefore,  $|f^n(A)| > \kappa_1$ , where  $\kappa_1 > 0$  is independent of  $A, n$  and  $t$  (for  $t$  small enough). Hence there exists a number  $N$  (independent of  $A, n$  and  $t$ ) such that  $f^{n+k}(A) \cap U_0 \neq \emptyset$  for some (minimal)  $k \leq N$ . In fact,  $U_0 \subset f^{n+k}(A)$ , because the boundary points of  $f^n(A)$  never return to  $U_0$  under iteration of  $f$ . Also,  $f^{n+k}: A \rightarrow f^{n+k}(A)$  is a diffeomorphism and  $A \setminus E_{n+k}$  is a subinterval of  $A$  such that  $f^{n+k}(A \setminus E_{n+k}) = U_1$ . The distortion of  $f^{n+k}$  is bounded by  $C'$  on  $A$ , by Lemma 3.9. Consequently

$$\frac{m(A \setminus E_{n+k})}{m(A)} \geq C'^{-1} \frac{|U_1|}{|I|}.$$

Hence there exists  $\gamma \in (0, 1)$ , independent of  $A, n, t$ , for which

$$m(A \cap E_{n+N}) \leq m(A \cap E_{n+k}) \leq \gamma m(A).$$

Summing over all connected components of  $E_n$ , we obtain  $m(E_{n+N}) \leq \gamma m(E_n)$ . The result for  $m(E_n)$  follows by induction. Since  $f(R_n) \subset E_{n-1}$  and  $f$  has a quadratic critical point,  $m(R_n) \lesssim \sqrt{m(E_{n-1})}$ , so we also obtain the result for  $m(R_n)$ .  $\square$

Denote  $f_t^{n+1}(0)$  by  $\xi_n(t)$ . The proof of the following lemma is based on [39]; the ideas go back at least to [8].

**Lemma 3.16.** *If  $\{f_t\}$  is transversal, there exist  $r_0 > 0, m_0 \geq 1$  and a sequence of positive numbers  $\gamma_n, n \geq m_0$  with*

- (a)  $\gamma_n/\gamma_{n+1} \simeq 1, \quad \lim_{n \rightarrow \infty} \gamma_n = 0;$
- (b)  $\gamma_n^{-1} \simeq |D\xi_n(0)| \simeq |Df_0^n(f_0(0))|;$
- (c)  $|\xi_n(\gamma_n) - \xi_n(0)| \geq r_0;$
- (d) *for all  $m_0 \leq k \leq n$ , the map  $\xi_k$  is monotone on  $[0, \gamma_n]$  and has a distortion bound*

$$\left| \log \frac{|D\xi_k(s)|}{|D\xi_k(t)|} \right| \leq 1 \quad \text{for all } s, t \in [0, \gamma_n];$$

(e)

$$\left| \log \frac{|Df_0^n(f_0(0))|}{|Df_t^n(f_t(0))|} \right| \leq 1 \quad \text{for all } t \in [0, \gamma_n].$$

**Proof.** Recall from Lemma 3.7 that  $|Df_0^k(f_0(0))| \geq C_0 \lambda^k$ . We use Tsujii [39] and only treat large  $n$ . From [39, Equation 3.3],

$$|Df_0^n(f_0(0))|^{-1} \simeq a^+(f_0(0), n; 0),$$

where  $a^+(x, n; t) = \left( 4e\kappa_1^2 \sum_{j=0}^{n-1} \frac{|Df_t^j(x)|}{|Df_t^j(f_t^j(x))|} \right)^{-1}$  and  $\kappa_1 > 1$ .

We choose  $\gamma_n$  equal to  $\gamma^{(\mu)}(0, n)$  in [39, Section 5]. By [39, Lemma 5.2] and the preceding Remarks with  $t = 0$ ,

- $|D\xi_n(0)| \simeq |Df_0^n(f_0(0))|;$
- $\gamma_n < |D\xi_n(0)|^{-1};$
- $\gamma_n \gtrsim a^+(f_0(0), n; 0).$

Hence we obtain (b) which in turn implies (a).

Bounds (d) and (e) correspond to [39,  $\Gamma 1$  and  $\Gamma 2$ ]. Finally, (c) follows from  $\gamma_n \simeq |D\xi_n(0)|^{-1}$  and (d).  $\square$

#### 4. First return maps

We continue to suppose that  $\{f_t\}$  is a Misiurewicz-rooted unimodal family. Let  $\Lambda_0$  be the closure of the post-critical orbit of  $f_0$ . Let  $\Lambda_t$  be its continuation, see Definition 3.13.

Where appropriate, we shall suppress the dependence on  $t$  from notation for better legibility.

Given the intervals  $U_j$ , as in Lemma 3.14, we denote by  $\phi_j: U_j \rightarrow U_j$  the first return map under iteration by  $f_t$ , and by  $\psi_j: I \rightarrow U_j$  the first entry map.

**Lemma 4.1.** *There are constants  $C > 1$ ,  $\theta_1 \in (0, \theta_0)$  such that for  $\theta \in (0, \theta_1)$ , if  $U_j$ ,  $j = 0, 1$ , are given by Lemma 3.14, if  $t$  is small and if  $x \in U_j$  with  $|x| > Ct$ , then*

$$|D\phi_j(x)| \geq 1000.$$

**Proof.** Let  $\delta_0 = \frac{1}{4} \text{dist}(\Lambda_0, 0)$ . Set  $y_0 = f_0(0) \in \Lambda_0$  and let  $y_t$  denote the continuation of  $y_0$ . Suppose that  $x$  is small and  $f_t(x) \neq y_t$ . Then

$$\begin{aligned} |f_t(x) - y_t| &\leq |f_t(x) - f_t(0)| + |f_t(0) - y_0| + |y_0 - y_t| \\ &\lesssim x^2 + t. \end{aligned}$$

Let  $W = (f_t(x), y_t)$  and set

$$n = \inf\{k \geq 0: |f_t^k(W)| \geq \delta_0\}.$$

As  $y_t$  is in the  $f_t$ -invariant set  $\Lambda_t$ ,

$$f_t^k(W) \cap (-\delta_0, \delta_0) = \emptyset \tag{4.1}$$

for all  $0 \leq k < n$ . By Lemma 3.8,  $|Df_t^k| \geq C\lambda^k$  on  $W$  for some  $C' > 0$  and  $\lambda > 1$  independent of  $x$  and  $t$ , for all  $0 \leq k < n$ . Hence  $n$  is finite. By Lemma 3.9,  $f_t^n$  has bounded distortion on  $W$ , independent of  $x$  and  $t$ . Hence,

$$|Df_t^n(f_t(x))| \gtrsim \frac{1}{|f_t(x) - y_t|}$$

and

$$|Df_t^{n+1}(x)| \gtrsim \frac{|x|}{|f_t(x) - y_t|} \gtrsim \frac{|x|}{x^2 + t}.$$

By Lemma 3.8, there is  $C' > 0$  such that the first entry map to any sufficiently small neighbourhood  $U$  of 0 has derivative at least  $C'$ . Further, if  $U \subset (-\delta_0, \delta_0)$ , then by (4.1), the first return time of  $x$  to  $U$  under  $f_t$  is at least  $n$ . Hence if  $\phi$  is the first return map to  $U$ , then, provided that  $U$  and  $t$  are small enough,

$$|D\phi(x)| \gtrsim \frac{|x|}{x^2 + t},$$

with the implied constant independent of  $U$  or  $t$ .

One can therefore choose  $C > 1$  and  $\theta_1 \in (0, \theta_0)$  so that, if  $U \subset (-\theta_1, \theta_1)$  and  $x \in U$ ,  $|x| \geq Ct$ , then  $|D\phi(x)| \geq 1000$ .

Now if  $\theta \in (0, \theta_1)$  and  $U_j$  are given by Lemma 3.14 with first return maps  $\phi_j$ , then the above estimates imply that  $|D\phi_j(x)| \geq 1000$  provided  $t$  is small enough,  $|x| \geq Ct$  and  $x \in U_j$ .  $\square$

**Lemma 4.2.** *If  $W$  is a branch of  $\psi_1$ , then there is an open interval  $\hat{W}$ , with  $W \subset \hat{W}$ , mapped diffeomorphically by  $f_t^n$  onto  $U_0$ , where  $\psi_1 = f_t^n$  on  $W$ .*

**Proof.** If  $W \ni 0$ , then  $n = 0$ ,  $f_t^n$  is the identity map and the claim is trivial indeed. Suppose  $0 \notin W$ . Let  $\hat{W} \supset W$  be the maximal open interval with  $f_t^n(\hat{W}) \subset U_0$ . Since  $f_t^k(\partial U_1) \cap U_0 = \emptyset$  for  $k \geq 1$ ,

$$f_t^j(\hat{W}) \cap \partial U_1 = \emptyset \quad \text{for all } 0 \leq j < n.$$

Since  $n$  is the first entry time on  $W$ ,

$$f_t^j(\hat{W}) \cap U_1 = \emptyset \quad \text{for all } 0 \leq j < n.$$

Hence  $f_t^n$  has no critical point in a neighbourhood of  $\overline{\hat{W}}$ , and maximality gives surjectivity.  $\square$

If  $\phi_1$  has a critical point, it is unique and equal to 0. Otherwise,  $\phi_1$  is not defined at 0. A branch of  $\phi_1$  containing 0 is called *central*.

**Lemma 4.3.** *Suppose that either 0 never returns to  $U_0$  or the first return of 0 to  $U_0$  lies in  $U_1$ . Let  $W$  be a non-central branch of  $\phi_1$ . Then there is an open interval  $\hat{W}$ , with  $W \subset \hat{W} \subset U_1$ , mapped diffeomorphically by  $f_t^n$  onto  $U_0$ , where  $\phi_1 = f_t^n$  on  $W$ . In case  $\phi_1$  has a central branch,  $\hat{W}$  is disjoint from it. On the non-central branches,  $|D\phi_1| \geq 5$ .*

**Proof.** As in the proof of Lemma 4.2, let  $\hat{W} \supset W$  be the maximal open interval with  $f_t^n(\hat{W}) \subset U_0$ . Then  $f_t^j(\hat{W}) \cap \partial U_1 = \emptyset$  for  $0 \leq j < n$ , in particular,  $\hat{W} \subset U_1$ . Since  $n$  is the first return time on  $W$ ,

$$f_t^j(\hat{W}) \cap U_1 = \emptyset \quad \text{for } 1 \leq j < n. \tag{4.2}$$

Therefore 0 is the only possible critical point of  $f_t^n$  on  $\hat{W}$ .

Next we show that  $0 \notin \hat{W}$ . Indeed, suppose that  $0 \in \hat{W}$ . Then by (4.2) and by the first return hypothesis,  $f_t^k(0) \notin U_0$  for  $1 \leq k < n$ , thus  $n$  is the first return time of 0 to  $U_0$ . Again by the first return hypothesis,  $f_t^n(0) \in U_1$ . Since 0 is the only critical point of  $f_t^n$  on  $\hat{W}$ , all points between 0 and  $W$  get mapped by  $f_t^n$  into  $U_1$ , so  $0 \in W$ , contradicting our assumption that  $W$  is non-central.

Since  $\hat{W}$  is the maximal open interval with  $f_t^n(\hat{W}) \subset U_0$  and  $f_t^n$  has no critical points on  $\hat{W}$ , it follows that  $f_t^n(\hat{W}) = U_0$ .

Now let us show that in case  $\phi_{1,t}$  has a central branch,  $\hat{W}$  is disjoint from it. Suppose that  $Z$  is the central branch with return time  $n_0$  and that  $\hat{W} \cap Z \neq \emptyset$ . Since  $0 \in Z$  and  $0 \notin \hat{W}$ , it follows that there is  $x \in \partial \hat{W} \cap Z$ . Then  $f_t^n(x) \in \partial U_0$ , so  $f_t^k(x) \notin U_0$  for all  $k \geq n$ , thus  $n_0 < n$ . Hence,  $f_t^n(\partial Z) \notin U_0$ , so  $\partial Z \cap \hat{W} = \emptyset$ . It follows that  $Z$  contains  $\hat{W}$  and  $n_0 = n$ , which contradicts  $n_0 < n$ .

Since  $\theta < \theta_0 < \frac{1}{10(1+\Delta)}$  and  $|U_1| < \theta \text{ dist}(U_1, \partial U_0)$ , the derivative estimate follows from Lemma 3.3.  $\square$

### 5. Breakdown of statistical stability

In this section, we suppose that our Misiurewicz-rooted unimodal family is transversal and prove Theorem 2.7. We again let  $\Lambda_0$  denote the closure of the post-critical orbit of  $f_0$  and  $\Lambda_t$  the continuation of  $\Lambda_0$ . The absolutely continuous invariant probability measure for  $f_0$  is  $\mu_0$ .

**Lemma 5.1.** *Given any  $\varepsilon > 0$ , there is a neighbourhood  $W_\Lambda$  of the post-critical set  $\Lambda_0$  of  $f_0$  and a  $C^\infty$  observable  $\varphi$  with  $\varphi \geq 0$  for which*

$$\varphi(x) = 1$$

for all  $x \in W_\Lambda$  and for which

$$\int \varphi d\mu_0 < \varepsilon.$$

**Proof.** By Lemma 3.15,  $m(\Lambda_0) = 0$ . As  $\Lambda_0$  is compact and  $\mu_0$  is absolutely continuous, Urysohn’s Lemma provides a continuous function which is 1 on  $\Lambda_0$  and 0 on a closed set of  $\mu_0$ -measure  $1 - \varepsilon$ . Perturbing this function, the result follows.  $\square$

Showing Theorem 2.7 therefore reduces to proving the following proposition, whose proof takes the rest of this section.

**Proposition 5.2.** *Let  $a > 0$ . There exists  $\alpha_0 > 0$  such that, for any neighbourhood  $W_\Lambda$  of  $\Lambda_0$  with the characteristic function  $1_{W_\Lambda}$ ,*

$$\limsup_{t \rightarrow 0^+} \int_I \bar{S}_{t, \lfloor at^{-1} \rfloor} 1_{W_\Lambda} dm \geq \alpha_0.$$

Our strategy is to construct a sequence  $t_n$  with  $\lim_{n \rightarrow \infty} t_n = 0$  such that: the maps  $f_{t_n}$  have 0 as a super-attracting periodic point; most of the *immediate basin of attraction* of the corresponding periodic orbit is contained in a small neighbourhood of  $\Lambda_0$ ; a definite proportion of all points in  $I$  enter the immediate basin in fewer than  $\lfloor t^{-1} \rfloor / 2$  iterates.

**Definition 5.3.** The *immediate basin of attraction* of a periodic point is the union of the connected components of the basin of attraction which contain points of the periodic orbit.

Let  $r_0, m_0, (\gamma_n)_{n \geq m_0}$  be as in Lemma 3.16. Let  $\theta_1 > 0$  be given by Lemma 4.1.

**Lemma 5.4.** *There are  $N \geq 1, \theta \in (0, \theta_1)$  and a sequence of parameters  $t_n > 0$  such that*

- (a)  $t_n \simeq \gamma_n \simeq |Df_{t_n}^n(f_{t_n}(0))|^{-1} \simeq |Df_0^n(f_0(0))|^{-1}$ ;
- (b) for some  $p_n \in [n, n + N]$ ,  $f_{t_n}^{p_n}(0) = 0$  and  $f_{t_n}^k(0) \notin (-\theta, \theta)$  for  $0 < k < p_n$ .

**Proof.** Recall that by Lemma 3.16, for  $m_0 \leq k \leq n$ , the map  $\xi_k$  is monotone on  $[0, \gamma_n]$  and has universally bounded distortion. Thus  $|\xi_k([0, \varepsilon \gamma_n])| \lesssim \varepsilon$  for  $\varepsilon > 0$ . For  $k < m_0$ , we bound  $|\xi_k([0, \varepsilon \gamma_n])| \leq \varepsilon \sup_{j < m_0} \sup_t |D\xi_j(t)|$ . Overall,

$$|\xi_k([0, \varepsilon \gamma_n])| \lesssim \varepsilon \quad \text{for all } k \leq n.$$

We choose  $\varepsilon_0$  small enough so that

$$\text{dist}(\xi_k([0, \varepsilon_0 \gamma_n]), 0) > \text{dist}(\Lambda_0, 0)/2 \quad \text{for all } k \leq n.$$

By Lemma 3.16,  $|\xi_n([0, \gamma_n])| \geq r_0$ ; since  $\xi_n$  has bounded distortion, there is an  $\varepsilon_1 > 0$  for which  $|\xi_n([0, \varepsilon_0 \gamma_n])| > \varepsilon_1$  for all large  $n$ . Note that  $\varepsilon_1 < \text{dist}(\Lambda_0, 0)/2$ . Fix  $N$  large so that, setting

$$Q_t = \bigcup_{k=1}^{N-1} f_t^{-k}(0),$$

$Q_0$  is  $\varepsilon_1/3$ -dense in  $I$ , see Lemma 3.11. For  $t$  small,  $Q_t$  is  $\varepsilon_1/2$ -dense. There is  $\theta \in (0, \theta_1)$  for which  $Q_t \cap (-\theta, \theta) = \emptyset$  for small  $t$ . Moreover  $\text{dist}(Q_t, \Lambda_0) \simeq 1$ .

Define

$$t_n = \min\{t \in [0, \gamma_n] : \xi_n(t) \in Q_t\}.$$

By construction,  $0 < t_n < \varepsilon_0 \gamma_n$  and (b) holds. By Lemma 3.16,  $\xi_n$  acts on  $[0, \gamma_n]$  as a diffeomorphism with bounded distortion. It follows from  $\text{dist}(\Lambda_0, Q_t) \simeq 1$  that  $|\xi_n(t_n) - \xi_n(0)| \simeq 1$ . Thus  $t_n \simeq \gamma_n$ ; the remaining relations in (a) follow from Lemma 3.16.  $\square$

We now work with the fixed map  $f = f_{t_n}$ , where  $n$  is as large as necessary. Write  $p = p_n$  for the period of 0. Let the intervals  $U_j$  be given by Lemma 3.14 for  $\theta$  from Lemma 5.4. Let  $\phi_1$  denote the first return map to  $U_1$ . An example graph of  $\phi_1$  is shown in Fig. 1.

Lemma 5.4 guarantees that the first return of 0 under  $f$  to  $U_0$  is  $0 \in U_1$ , thus by Lemma 4.3,  $\phi_1$  restricted to  $U_1$  has a unimodal central branch which we denote by  $Z$ ; all other branches are full with a uniform distortion bound. On  $Z$ ,  $\phi_1 = f^p$ . We denote the immediate basin of attraction (with respect to  $\phi_1$ ) of 0 by  $V$ . As  $\phi_1(0) = 0$ ,  $V$  is an interval.

**Lemma 5.5.** *Given any neighbourhood  $W_\Lambda$  of  $\Lambda_0$  and  $\varepsilon > 0$ , the following holds for all  $n$  large enough. For all  $x \in V$  and  $k \geq 1$ , the Birkhoff average of the characteristic function  $1_{W_\Lambda}$  of  $W_\Lambda$  satisfies*

$$\bar{S}_{t_n, k} 1_{W_\Lambda}(f(x)) \geq 1 - \varepsilon.$$

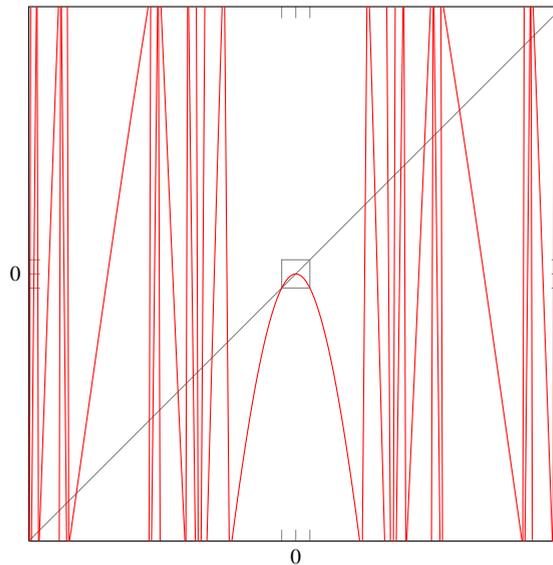


Fig. 1. Graph of  $\phi_1 : U_1 \rightarrow U_1$  when 0 is a periodic point. Between every two branches there are countably many other branches;  $\phi_1$  is uniformly expanding outside the small invariant interval in the middle.

**Proof.** Note that  $\phi_1(V) \subset V$  and recall that the first return of 0 to  $U_0$  is at time  $p$  with  $n + 1 \leq p \leq n + N$ . Given  $\varepsilon > 0$  we shall show, for large  $n$  and  $j \leq (1 - \varepsilon)n$ , that  $f^j(V)$  and  $\text{dist}(f^j(V), \Lambda_0)$  are sufficiently small to guarantee that  $f^j(V) \subset W_\Lambda$ . Since  $f^{p\ell}(V) \subset V$  for each  $\ell \geq 0$ , this implies that  $f^j(V) \subset W_\Lambda$  for all  $p\ell + 1 \leq j < p\ell + (1 - \varepsilon)n$ . From this, the Birkhoff estimate follows.

For  $j = 1, \dots, n$ ,  $f^j(V) \cap U_0 = \emptyset$ . Lemma 3.8 implies that  $|f^j(V)|$  is exponentially small in  $n - j$ . By Lemma 3.7,  $|Df_0^{n-j}(f_0^k(0))| \gtrsim \lambda^{n-j}$ . With the estimates of Lemma 3.16, one deduces that  $\text{dist}(f^j(0), \Lambda_0) = \text{dist}(\xi_{j-1}(t_n), \Lambda_0)$  is exponentially small in  $n - j$ . Thus so is  $\text{dist}(f^j(V), \Lambda_0)$ . The proof is complete.  $\square$

We establish properties of  $\phi_1$  on  $Z$ .

**Lemma 5.6.**

- (a)  $|Z| \simeq t_n^{1/2}$  and  $|V| \simeq t_n$ ;
- (b) there exists  $\eta > 0$ , independent of  $n$ , such that  $|D\phi_1| > e^{n\eta}$  on  $Z \setminus \phi_1^{-1}(Z)$ ;
- (c)  $\log |D\phi_1| > 1/2$  on  $U_1 \setminus V$ .

**Proof.** Since  $n + 1 \leq p \leq n + N$ ,  $|Df^n(f(0))| \simeq |Df^{p-1}(f(0))|$ . Let  $\psi_1$  be the first entry map to  $U_1$ . Its branches have bounded distortion (Lemma 4.2), so Lemma 5.4 entails that

$$|\psi_1'(f(0))| \simeq |Df^{p-1}(f(0))| \simeq |Df^n(f(0))| \simeq t_n^{-1}.$$

The interval  $V$  is a neighbourhood of the non-degenerate critical point,  $|f(V)| \simeq |V|^2$ . At the same time,  $|\phi_1(V)| \simeq |V|$  (see Fig. 1:  $Z$  is the domain of the central branch, and  $V$  is the small invariant interval in the middle). Observe that  $f^p = \psi_1 \circ f$  on  $Z$ . Hence

$$|V| \simeq |\psi_1'(f(0))|^{-1} \simeq t_n.$$

Meanwhile,  $|\phi_1(Z)| \simeq 1$ , so  $|f(Z)| \simeq t_n$  and  $|Z| \simeq t_n^{1/2}$ . This proves (a).

Let  $I_1 = Z \cap \phi_1^{-1}(Z)$ . By a similar argument,

$$|I_1| \simeq \sqrt{t_n^{1/2} t_n} = t_n^{3/4}.$$

Let  $J_0$  be the union of the pair of symmetric intervals  $Z \setminus I_1$ , then  $\text{dist}(J_0, 0) \simeq t_n^{3/4}$  (non-degeneracy implies  $I_1$  is roughly centred on 0). On  $J_0$ ,

$$|Df| \gtrsim t_n^{3/4}$$

so, on the same set,

$$|D\phi_1| \gtrsim t_n^{-1} t_n^{3/4} = t_n^{-1/4}.$$

By Lemma 5.4,  $t_n^{-1} \simeq |Df_0^n(f_0(0))|$ , and exponential growth of the latter implies the existence of an  $\eta > 0$  for which  $t_n < \exp(-5n\eta)$  (for all  $n$ ). Combined with the previous sentence, we obtain (b).

It remains to prove (c). On  $I_1$ , we claim

$$\frac{D\phi_1(x)}{a_0 x D\psi_1(f(0))} = 1 + o(1) \quad \text{as } n \rightarrow \infty, \tag{5.1}$$

where  $a_0 = D^2 f_0(0) \neq 0$ . Observe that  $\phi_1 = \psi_1 \circ f$  and  $f(I_1)$  is mapped by  $\psi_1$  into  $Z$ . We have  $|Z| < t_n^{1/3} \text{dist}(Z, \partial U_1)$ ; by the Koebe Principle,  $\psi_1 : f(I_1) \rightarrow Z$  has distortion bounded by  $1 + o(1)$ . By continuity,  $D^2 f(x) = a_0(1 + o(1))$ . Integrating,  $Df(x) = a_0 x(1 + o(1))$ , which gives the claim.

This time, integrate  $D\phi_1$  to get

$$\phi_1(x) = \frac{bx^2}{2}(1 + o(1))$$

with  $b = a_0 D\psi_1(f(0))$ . The fixed point  $y$  in  $\partial V$  satisfies

$$|y| = |\phi_1(y)| = \frac{|b|y^2}{2}(1 + o(1)),$$

so  $|y| = \frac{2}{|b|}(1 + o(1))$ . Inserting this in (5.1) gives  $|D\phi_1(y)|/2 = 1 + o(1)$  and

$$D\phi_1(x) = \frac{x}{y} D\phi_1(y)(1 + o(1)).$$

If  $y'$  is the other boundary point of  $V$ , then  $y' = -y(1 + o(1))$ . Hence on  $Z \setminus V$ ,  $|x/y| \geq 1 + o(1)$  and  $\log |D\phi_1(x)| > \log 2 - 1/10 > 1/2$ .  $\square$

Let  $\chi : U_1 \setminus V \rightarrow U_1 \setminus Z$  be the first entry map to  $U_1 \setminus Z$ . By Lemma 5.6, it is well-defined (almost surely). On  $U_1 \setminus Z$  it is identity, while on  $Z \setminus V$  it has countably many branches, each being mapped diffeomorphically onto a connected component of  $U_1 \setminus Z$ .

We define  $F : U_1 \rightarrow U_1$  by

$$F(x) = \begin{cases} \phi_1 \circ \chi(x), & x \in U_1 \setminus V, \\ A(x), & x \in V, \end{cases} \tag{5.2}$$

where  $A$  is an affine homeomorphism between  $V$  and  $U_1$ . Let  $\tau : U_1 \setminus V \rightarrow \mathbb{N}$  be the corresponding inducing time, so  $F(x) = f^{\tau(x)}(x)$ , and set  $\tau = 1$  on  $V$ .

**Lemma 5.7.**

- (a) All branches of all iterates of  $F$  have uniformly bounded distortion (independent of the iterate and of  $n$ ). The image of such a branch is  $U_1$ .
- (b) There exists a constant  $\alpha > 0$ , independent of  $n$ , so that

$$m(\tau = j) \lesssim \exp(-\alpha\sqrt{j}) \quad \text{for all } j.$$

**Proof.** To prove (a), it is enough to show that branches of  $F$  other than  $V$  are mapped onto  $U_1$  and are  $\Delta$ -extensible, with extension contained in  $U_1 \setminus V$ . Let us do this. By Lemma 4.3, this holds for branches of  $\phi_1$  contained in  $U_1 \setminus Z$ . Each branch of  $\chi$  is mapped diffeomorphically by  $\chi$  onto a connected component of  $U_1 \setminus Z$  and (a) follows.

Now we prove (b). Set  $I_0 := Z$  and, inductively,

$$I_{k+1} := \phi_1^{-1}(I_k) \cap Z.$$

These are nested intervals whose intersection (over all  $k$ ) is  $V$ . Denote by  $J_k$  the pair of symmetric intervals  $I_k \setminus I_{k+1}$ . On each  $J_k$ ,  $\chi = \phi_1^{k+1} = f^{p(k+1)}$ . By Lemma 5.6,  $\log |D\phi_1| \gtrsim n$  on  $J_0$  and  $\log |D\phi_1| > 1/2$  on  $J_k$ . Thus with some  $\alpha' > 0$ , on  $J_k$ ,

$$|D\chi| \gtrsim \exp(\alpha'(n+k)).$$

If we take  $\alpha'$  small enough, we also have, by Lemma 3.15,

$$m(\{x \in U_1 \setminus Z : \tau(x) = j\}) \lesssim \exp(-\alpha'j).$$

Since  $F = \phi_1 \circ \chi = \phi_1 \circ f^{p(k+1)}$  on  $J_k$ , then

$$m(\{x \in J_k : \tau(x) = j\}) = 0 \quad \text{if } k \geq \lfloor j/p \rfloor$$

while, if  $k \leq \lfloor \frac{j}{p} \rfloor - 1$ ,

$$m(\{x \in J_k : \tau(x) = j\}) \lesssim \exp(-\alpha'(n+k) - \alpha'(j - p(k+1))).$$

Observe that, letting  $k$  go from 1 to  $\lfloor j/p \rfloor - 1$ , the above forms a geometric sequence in  $k$  with ratio  $\exp(\alpha'(p-1))$ . Its sum is approximated (to within a multiplicative constant) by its maximal term, that is with  $k = \lfloor j/p \rfloor - 1$ . Hence

$$\begin{aligned} m(\{x \in Z \setminus V : \tau(x) = j\}) &= \sum_{k \geq 0} m(\{x \in J_k : \tau(x) = j\}) \\ &= \sum_{k=0}^{\lfloor j/p \rfloor - 1} m(\{x \in J_k : \tau(x) = j\}) \\ &\lesssim \exp(-\alpha'(n + j/p)) \\ &\lesssim \exp(-\alpha'(n + j/n)) \leq \exp(-2\alpha'\sqrt{j}), \end{aligned}$$

using  $n + 1 \leq p \leq n + N$  to pass to the last line. This proves (b).  $\square$

**Lemma 5.8.** *For every  $C > 0$ , there is  $\delta > 0$  such that for all sufficiently large  $n$ ,*

$$m(\{x \in U_1 : f^k(x) \in V \text{ for some } k \leq Ct_n^{-1}\}) \geq \delta.$$

**Proof.** We redefine  $f$  on  $V = V_n$  so that  $f : V \rightarrow U_1$  is the affine homeomorphism  $A$  as in (5.2). This does not change when a point first enters  $V$  (noting  $k = 0$  is possible) and does not change  $F$ . With this modification,  $F$  is the induced map for  $f$  with inducing time  $\tau$ . Let  $\tau_k = \sum_{j=0}^{k-1} \tau \circ F^j$ .

Let  $\nu$  be the Lebesgue measure on  $U_1$ , normalized so that  $\nu(U_1) = 1$ . Let

$$\begin{aligned} W_k &= \{x \in U_1 : f^j(x) \notin V \text{ for all } j \leq k\}, \\ W'_k &= \{x \in U_1 : f^j(x) \notin V \text{ for all } j \leq \tau_k\}. \end{aligned}$$

By Lemma 5.7, all branches of  $F$  are full and have universally bounded distortion. Consequently, the set of points not entering  $V$  in  $k$  iterates of  $F$  is exponentially small, namely

$$\nu(W'_k) \leq (1 - C_1|V|)^k,$$

where  $C_1$  is a universal constant. Now,  $W_k \subset W'_\ell \cup \{\tau_\ell > k\}$  for all  $\ell \geq 0$ . Hence

$$\begin{aligned} \nu(W_k) &\leq \nu(W'_\ell) + \nu(\{\tau_\ell > k\}) \\ &\leq (1 - C_1|V|)^\ell + \nu(\{\tau_\ell > k\}). \end{aligned}$$

We claim that there exists a constant  $c > 0$  such that  $\nu(\{\tau_{ck} > k\}) \rightarrow 0$  as  $k \rightarrow \infty$ , uniformly in  $n$ . Suppose that the claim is true. Setting  $k = Ct_n^{-1}$  and  $\ell = ck$ , and using  $|V| \simeq t_n$ , we obtain

$$v(W_{Ct_n^{-1}}) \leq (1 - at_n)^{bt_n^{-1}} + o(1) = e^{-ab} + o(1),$$

with some  $a, b > 0$ . This implies the result.

It remains to verify the claim. The map  $F : U_1 \rightarrow U_1$  is Gibbs-Markov with full images. By Lemmas 3.8 and 5.7, the expansion and distortion bounds of  $F$  can be chosen independent of  $n$ . Let  $\mu$  be the  $F$ -invariant absolutely continuous probability measure on  $U_1$ , and let  $\bar{\tau} = \int \tau d\mu$ . Observe that  $\tau$  is constant on the branches of  $F$ , and by Lemma 3.15,  $|\tau|_{L^2(\mu)} \simeq 1$ . It is standard (see Lemma A.2) that

$$|\tau_k - k\bar{\tau}|_{L^2(\mu)} \lesssim k^{-1/2}.$$

It is also standard that  $d\mu/dv \simeq 1$ , so  $|\tau_k - k\bar{\tau}|_{L^2(v)} \lesssim k^{-1/2}$ , which implies the claim.  $\square$

Let  $a > 0$  and take  $C < a/2$ . By the preceding lemma, there is a set of measure  $\delta > 0$  of points which enter  $V$  in fewer than  $at^{-1}/2$  iterates. Applying Lemma 5.5,  $\bar{S}_{t_n, at_n^{-1}} 1_{W_\Lambda}(x) \geq (1 - \varepsilon)/2$  for every  $x$  in this set, provided  $n$  is large enough. This proves Proposition 5.2 with  $\alpha_0 = \delta(1 - \varepsilon)/2$ .

### 6. Persistence of statistical stability

In this section we prove Theorem 2.6. Our strategy is as follows:

- [Proposition 6.2 and §6.2] We construct a particular inducing scheme for  $f_t$ , which we use to approximate  $f_t$  with a nonuniformly expanding map  $\hat{f}_t$  which admits an absolutely continuous invariant probability measure  $\hat{\mu}_t$ . The construction is such that  $\hat{f}_0 = f_0$  and  $\hat{\mu}_0 = \mu_0$ . The map  $\hat{f}_t$  has uniform in  $t$  bounds on return times, expansion and distortion. Further,  $\hat{f}_t$  agrees with  $f_t$  everywhere except on a set of Lebesgue measure of order  $t$ .
- [Lemma 6.6] Suppose that  $\varphi : I \rightarrow \mathbb{R}$  is Lipschitz. We show that for all  $n \geq 1$ ,

$$\int_I \left| \frac{1}{n} \sum_{j=0}^{n-1} \varphi \circ \hat{f}_t^j - \int \varphi d\hat{\mu}_t \right| dm \leq Cn^{-1/2} |\varphi|_{\text{Lip}}, \tag{6.1}$$

where the constant  $C$  does not depend on  $t$  and  $|\cdot|_{\text{Lip}}$  is the Lipschitz norm,

$$|\varphi|_{\text{Lip}} = \sup_{x \in I} |\varphi(x)| + \sup_{x \neq y \in I} \frac{|\varphi(x) - \varphi(y)|}{|x - y|}.$$

- [Lemma 6.7] We show that  $f_t$  agrees with  $\hat{f}_t$  on time horizons smaller than  $t^{-1}$ , namely that if  $n(t) = o(t^{-1})$ , then

$$\lim_{t \rightarrow 0} m\{x \in I : f_t^j(x) = \hat{f}_t^j(x) \text{ for all } j \leq n(t)\} = 1.$$

For a bounded observable  $\varphi : I \rightarrow \mathbb{R}$ , this naturally implies that

$$\lim_{t \rightarrow 0} \int_I \left| \frac{1}{n(t)} \sum_{j=0}^{n(t)-1} \varphi \circ \hat{f}_t^j - \frac{1}{n(t)} \sum_{j=0}^{n(t)-1} \varphi \circ f_t^j \right| dm = 0. \tag{6.2}$$

- [Lemma 6.8] Using continuity of the map  $(x, t) \mapsto f_t(x)$  and (6.1), we prove that

$$\int \varphi d\hat{\mu}_t \rightarrow \int \varphi d\mu_0 \quad \text{as } t \rightarrow 0. \tag{6.3}$$

From this point, all is straightforward. By (6.1) and (6.3), if  $n(t) \rightarrow \infty$  as  $t \rightarrow 0$ ,

$$\lim_{t \rightarrow 0} \int_I \left| \frac{1}{n(t)} \sum_{j=0}^{n(t)-1} \varphi \circ \hat{f}_t^j - \int \varphi d\mu_0 \right| dm = 0.$$

Combining this with (6.2), we obtain that for all Lipschitz  $\varphi : I \rightarrow \mathbb{R}$  and  $n(t)$  with  $\lim_{t \rightarrow 0} n(t) = \infty$  and  $n(t) = o(t^{-1})$ ,

$$\lim_{t \rightarrow 0} \int_I \left| \frac{1}{n(t)} \sum_{j=0}^{n(t)-1} \varphi \circ f_t^j - \int \varphi d\mu_0 \right| dm = 0.$$

This gives the result of Theorem 2.6 for Lipschitz observables. Generalisation to the class of continuous observables is automatic: every continuous observable can be arbitrarily well approximated by a Lipschitz observable in the uniform topology.

In the rest of this section we implement the strategy above. Where there is no ambiguity, we suppress the dependence on  $t$ .

**Remark 6.1.** One of the main difficulties in our proof is the construction of the approximating map  $\hat{f}_t$ , which allows a suitable inducing scheme and which coincides with  $f_t$  everywhere except on a set of Lebesgue measure of order  $t$ .

The proof for time horizons of order  $o(t^{-1/2})$  can be made significantly simpler than that for  $o(t)$ , as we only have to avoid a set of size  $t^{-1/2}$ . In this case,  $\hat{f}_t$  can be taken to be equal to  $f_t$  everywhere outside the central branch  $Z$  of the first return map to  $U_0$ , if such a branch exists; on  $Z$ , we can define  $\hat{f}_t$  as an affine bijection between  $Z$  and  $U_1$ .

Then it can be verified, using results of §3 and §4, that the first return map to  $U_1$  under  $\hat{f}_t$  has all of its branches full with universally bounded distortion, and a uniform in  $t$  exponential bound on return times. Similarly to Lemma 5.6, one can show that  $|Z| \lesssim t^{1/2}$ . With this, the strategy above works, rendering unnecessary most of §6.1.

### 6.1. Inducing scheme

Recall that  $\phi_1 : U_1 \rightarrow U_1$  is the first return map under  $f$ . It is constructed to have countably many branches, and all non-central branches (i.e. not containing 0) are mapped by  $\phi_1$  to  $U_1$  diffeomorphically.

Let  $V = (-Ct, Ct)$ , where  $C$  is the constant from Lemma 4.1. Then  $|D\phi_1| > 1000$  on  $U_1 \setminus V$ .

**Proposition 6.2.** *For small enough  $t$ , there exists a partition  $\mathcal{P}$  of  $U_1$  into open intervals, modulo a zero measure set. Each interval  $J \in \mathcal{P}$  is coloured blue or red, and there is a function  $\rho : U_1 \rightarrow \mathbb{N} \cup \{0\}$ , constant on each  $J$  with value  $\rho(J)$ , such that:*

- (a) if  $J$  is red, then  $f^{\rho(J)}(J) \subset V$ ;
- (b) if  $J$  is blue, then  $\rho(J) > 0$  and  $f^{\rho(J)} : J \rightarrow U_1$  is a diffeomorphism with universally bounded distortion;
- (c)  $m(\cup\{J \in \mathcal{P} : J \text{ is red}\}) \lesssim t$ ;
- (d)  $\int_{U_1} \rho^2 dm \simeq 1$ .

The proof of Proposition 6.2 takes the rest of this subsection. To simplify notation, if  $W$  is a branch of  $\phi_1$  intersecting  $\partial V$ , we consider the connected components of  $W \setminus \partial V$  as separate branches of  $\phi_1$ . In particular, if  $W'$  is a branch of  $\phi_1^k$ , then  $\phi_1^p(W') \cap \partial V = \emptyset$  for  $0 \leq p < k$ .

Let  $\tau : U_1 \rightarrow \mathbb{N}$  be the first return time,

$$\tau(x) = \inf\{k \geq 1 : f^k(x) \in U_1\},$$

so  $\phi_1 = f^\tau$ . Let  $\tau_k = \sum_{j=0}^{k-1} \tau \circ \phi_1^j$ . Note that if  $W$  is a branch of  $\phi_1^k$ , then as a consequence of Lemma 3.14(c), for each  $j \leq \tau_k(W)$  either  $f^j(W) \subset U_0$  or  $f^j(W) \cap U_0 = \emptyset$ .

We construct a nested sequence of partitions  $\mathcal{P}_k$ ,  $k \geq 0$ , of  $U_1$  into open intervals. To each interval we assign a colour (yellow, blue or red), an index and a height (integers). Let  $\mathcal{P}_0 = \{U_1\}$  be the trivial partition. We set the height of its only element to 0, index to 0 and colour it yellow. For  $k \geq 1$ , we construct  $\mathcal{P}_k$  as a refinement of  $\mathcal{P}_{k-1}$  inductively:

- We leave the blue and red intervals intact, with the same height and index.
- We partition each yellow  $J \in \mathcal{P}_{k-1}$  into the branches of the map  $\phi_1^k : J \rightarrow U_1$ . For each such new element  $W$  of  $\mathcal{P}_k$ :
  - If  $\phi_1^{k-1}(W) \subset V$ , then we colour  $W$  red. Otherwise,  $\phi_1^{k-1}(W) \cap V = \emptyset$ . If  $\phi_1^k : W \rightarrow U_1$  is a  $U_0$ -extensible diffeomorphism, we colour  $W$  blue. Otherwise we colour  $W$  yellow.

– We set

$$\text{height}(W) = \begin{cases} k - 1, & W \text{ is red} \\ k, & \text{otherwise} \end{cases}$$

and

$$\text{index}(W) = \#\{0 < j \leq \tau_{\text{height}(W)}(W) : f^j(W) \subset U_0\}.$$

**Lemma 6.3.** For all  $\ell \geq 0$ ,

- $\sum_{k \geq 0} \#\{J \in \mathcal{P}_k : J \text{ is yellow with index } \ell\} \leq 6^\ell.$
- $\sup_{k \geq 0} \#\{J \in \mathcal{P}_k : J \text{ is red with index } \ell\} \leq 6^\ell.$

**Proof.** Suppose that  $J \in \mathcal{P}_{k-1}$  is yellow with index  $\ell$ . In  $\mathcal{P}_k$  it is partitioned into subintervals. We claim that among these:

- (a) there is at most 1 red interval, its index is  $\ell$ ;
- (b) all yellow intervals have index at least  $\ell + 1$ , and there are at most 4 of them with index  $\ell + j$  for each  $j \geq 1$ .

A recursive estimate then implies that the number of yellow intervals contributing to the above sum is bounded by  $6^\ell$ . The same estimate holds then for red intervals and the result follows. We justify the claim now.

To each branch of  $\phi_1^k$  contained in  $J$  corresponds a branch of the restriction  $\phi_1 : \phi_1^{k-1}(J) \rightarrow U_1$ . The red interval corresponds to  $V$  intersected with  $\phi_1^{k-1}(J)$ . The statement of (a) is immediate.

Let  $\hat{J}$  be a connected component of  $\phi_1^{k-1}(J) \setminus V$ . Let  $W$  be a branch of the restriction  $\phi_1 : \hat{J} \rightarrow U_1$  with  $\tau = n$  on  $W$ . To  $W$  corresponds the element  $\hat{W} := \phi_1^{-(k-1)}(W) \cap J$  of  $\mathcal{P}_k$ , which is yellow or blue.

We call  $W$  *unobstructed* if  $f^n : W \rightarrow U_1$  is a diffeomorphism and there is an open interval  $W_0 \subset \hat{J}$ , compactly containing  $W$ , such that  $f^n : W_0 \rightarrow U_0$  is a diffeomorphism. Otherwise  $W$  is obstructed. Note that obstruction depends on  $\hat{J}$  and that  $\hat{W}$  can only be yellow if  $W$  is obstructed.

Let us examine the case when  $W$  is obstructed. There are  $w \in \partial W$  and  $v \in \overline{\hat{J}} \setminus W$  with  $[w, v] \cap W = \emptyset$  such that (noting  $v$  and  $w$  may coincide)

- $f^n$  is monotone on  $W \cup [w, v]$ ,
- $f^n([w, v])$  does not contain a connected component of  $U_0 \setminus U_1$ ,
- either  $Df^n(v) = 0$  or  $v \in \partial \hat{J}$ .

Since  $f^n([w, v])$  does not contain a connected component of  $U_0 \setminus U_1$ , it follows (via Lemma 3.14(c)) that  $f^p([w, v])$  does not contain a point of  $\partial U_0 \cup \partial U_1$  for all  $0 \leq p < n$ . This implies that  $f^p([w, v]) \cap U_1 = \emptyset$  for all  $0 < p < n$ . Therefore  $Df^n(v) \neq 0$ , so  $v \in \partial \hat{J}$ .

As  $f^n$  is monotone on  $W \cup [w, v]$ , there is a one-to-one correspondence between obstructed branches  $W \subset \hat{J}$  of  $\phi_1$  and a subset of the set of pairs  $(v, n) \in \partial \hat{J} \times \mathbb{N}$  for which  $f^n(v) \in U_0$ . For each such  $W$  and associated  $(v, n)$ , there is a unique  $j(v, n) := \#\{0 \leq p \leq n : f^p(v) \in U_0\}$ . Moreover, for  $0 \leq p \leq n$ , either  $f^p(W \cup [w, v]) \subset U_0$  or  $f^p(W \cup [w, v]) \cap U_0 = \emptyset$ , from which it follows that

$$j(v, n) = \#\{0 \leq p \leq n : f^p(W) \subset U_0\}.$$

Hence to each yellow element  $\hat{W} \subset J$  in  $\mathcal{P}_k$ , there is a unique obstructed branch  $W$  with associated  $\hat{J}$  and pair  $(v, n)$ . The index of  $\hat{W}$  is  $\ell + j(v, n)$ . With at most two ways to choose  $\hat{J}$  as a connected component of  $\phi_1^{k-1}(J) \setminus V$ , and two possibilities for  $v \in \partial \hat{J}$ , the claim and (b) follow.  $\square$

**Lemma 6.4.**

$$\sup_{n \geq 0} \sum_{\substack{J \in \mathcal{P}_n, \\ J \text{ is red}}} |J| \lesssim t \quad \text{and} \quad \sum_{n \geq 0} \sum_{\substack{J \in \mathcal{P}_n, \\ J \text{ is yellow}}} |J| \lesssim 1.$$

**Proof.** Suppose that  $J \in \mathcal{P}_n$  is an interval with index  $\ell$  and height  $h$ . By Lemma 4.1, the first return map to  $U_0$ , restricted to  $U_0 \setminus V$ , is expanding by a factor of at least 1000. By construction,  $\phi_1^k(J)$  does not intersect  $V$  for  $k < h$ . Thus:

- $|D\phi_1^h| \geq 1000^\ell$  on  $J$ , and  $\phi_1^h(J) \subset U_1$ , so

$$|J| \lesssim 1000^{-\ell};$$

- moreover, if  $J$  is red, then  $\phi_1^h(J) \subset V$ , so

$$|J| \lesssim 1000^{-\ell} |V| \lesssim 1000^{-\ell} t.$$

By Lemma 6.3,  $\mathcal{P}_n$  has at most  $6^\ell$  red intervals of index  $\ell$ , thus

$$\sum_{\substack{J \in \mathcal{P}_n, \\ J \text{ is red}}} |J| \lesssim \sum_{\ell \geq 0} 6^\ell 1000^{-\ell} t \lesssim t.$$

The result for red intervals follows. The argument for yellow intervals is similar.  $\square$

Let  $\mathcal{P} = \bigvee_n \mathcal{P}_n$ . By Lemma 6.4,  $\mathcal{P}$  is a partition of  $U_1$  into open intervals (blue and red), modulo a zero measure set. For  $J \in \mathcal{P}$ , let  $\rho(J) = \tau_{\text{height}(J)}$ . This defines  $\rho: U_1 \rightarrow \mathbb{N}$  with value  $\rho(J)$  on each  $J \in \mathcal{P}$ .

By construction,  $\rho$  satisfies (a), (b) and (c) of Proposition 6.2. It remains to prove (d).

**Lemma 6.5.**  $\int_{U_1} \rho^2 dm \simeq 1.$

**Proof.** It is clear that  $\int_{U_1} \rho^2 dm \gtrsim 1.$

Let  $J \in \mathcal{P}$ , so  $J$  is red or blue. Let  $h = \text{height}(J)$  and for  $k \leq h$ , let  $J_k$  be the element of  $\mathcal{P}_k$  containing  $J$ . Each  $J_k, k < h$ , is yellow, while  $J_h$  is yellow or blue. Then

$$\rho(J) = \sum_{k=0}^{h-1} \tau \circ \phi_1^k(J_{k+1}).$$

Define  $\rho_i$  at a point  $x$  by:  $\rho_i(x) = \tau \circ \phi_1^k(x)$  if  $x$  is contained in a yellow interval  $J' \in \mathcal{P}_k$  with index  $i$  and height  $k$ , for some  $k$ , but  $x$  is not contained in a red interval of height  $k$  (in  $\mathcal{P}_{k+1}$ ), and  $\rho_i(x) = 0$  otherwise. Then

$$\rho = \sum_{i=0}^{\infty} \rho_i.$$

Let  $J \in \bigcup_{n \geq 0} \mathcal{P}_n$  be yellow. The map  $\phi_1^{\text{height}(J)}: J \rightarrow U_1$  is monotone and, following the proof of Lemma 6.4, it is expanding by a factor of at least  $1000^{\text{index}(J)}$ . Using Lemma 3.15,

$$\int_J \tau^2 \circ \phi_1^{\text{height}(J)} dm \lesssim 1000^{-\text{index}(J)} \int_{U_1} \tau^2 dm \lesssim 1000^{-\text{index}(J)}.$$

Let  $i \geq 0$ . Let  $\mathcal{A}_i := \{J \in \bigcup_{n \geq 0} \mathcal{P}_n : J \text{ is yellow with index } i\}$ . By Lemma 6.3,  $\#\mathcal{A}_i \leq 7^i$ . Observe that

$$\rho_i = \sum_{J \in \mathcal{A}_i} \tau \circ \phi_1^{\text{height}(J)} \Big|_J.$$

The elements of  $\mathcal{A}_i$  are pairwise disjoint, thus

$$\int_{U_1} \rho_i^2 dm = \sum_{J \in \mathcal{A}_i} \int_J \tau^2 \circ \phi_1^{\text{height}(J)} dm \lesssim 7^i \cdot 1000^{-i} \leq 100^{-i}.$$

Finally,

$$\left[ \int_{U_1} \rho^2 dm \right]^{1/2} \lesssim \sum_{i=0}^{\infty} \left[ \int_{U_1} \rho_i^2 dm \right]^{1/2} \lesssim 1. \quad \square$$

### 6.2. Approximation with nonuniformly expanding map

Let  $\mathcal{P}$  be the partition given by Proposition 6.2. For an interval  $J \subset U_1$ , let  $\hat{f}_J : J \rightarrow U_1$  be a linear bijection. Define  $\hat{f} : I \rightarrow \mathbb{R}$  and  $\hat{\rho} : U_1 \rightarrow \mathbb{N}$ ,

$$\hat{f}(x) = \begin{cases} \hat{f}_J(x), & \text{if } x \in J, J \in \mathcal{P} \text{ is red,} \\ f(x), & \text{else,} \end{cases}$$

$$\hat{\rho}(x) = \begin{cases} 1, & \text{if } x \in J, J \in \mathcal{P} \text{ is red,} \\ \rho(x), & \text{else.} \end{cases}$$

Let  $\hat{F} : U_1 \rightarrow U_1$ ,  $\hat{F}(x) = \hat{f}^{\hat{\rho}(x)}(x)$ . In particular,  $\hat{F}$  coincides with  $f^\rho$  on all blue elements of  $\mathcal{P}$ . Our construction ensures that there are constants  $C > 0$  and  $\lambda > 1$ , independent of  $t$ , such that for every  $J \in \mathcal{P}$  and  $x, y \in J$ :

- the restriction  $\hat{F} : J \rightarrow U_1$  is a bijection;
- $|\hat{F}(x) - \hat{F}(y)| \geq \lambda|x - y|$ ;
- $|\log |D\hat{F}(x)| - \log |D\hat{F}(y)|| \leq C|\hat{F}(x) - \hat{F}(y)|$ ;
- $|\hat{f}^j(x) - \hat{f}^j(y)| \leq C|\hat{F}(x) - \hat{F}(y)|$  for all  $0 \leq j \leq \hat{\rho}(J)$ ;
- $\int_{U_1} \hat{\rho}^2 dm \leq C$ .

That is,  $\hat{f}$  is a *nonuniformly expanding map* as in Appendix A. There is a unique absolutely continuous  $\hat{f}$ -invariant probability measure  $\hat{\mu}$ .

**Lemma 6.6.** For all Lipschitz  $\varphi : I \rightarrow \mathbb{R}$  and  $n \geq 1$ ,

$$\int_I \left| \frac{1}{n} \sum_{j=0}^{n-1} \varphi \circ \hat{f}^j - \int \varphi d\hat{\mu} \right| dm \leq Cn^{-1/2} |\varphi|_{\text{Lip}},$$

where the constant  $C$  does not depend on  $t$ .

**Proof.** By Lemma A.3,

$$\int \left| \frac{1}{n} \sum_{j=0}^{n-1} \varphi \circ \hat{f}^j - \int \varphi d\hat{\mu} \right| d\hat{\mu} \lesssim n^{-1/2} |\varphi|_{\text{Lip}}. \tag{6.4}$$

Note that the integral is taken with respect to the invariant measure  $\hat{\mu}$  rather than  $m$ . It remains to establish an appropriate connection between  $m$  and  $\hat{\mu}$ . For this, we follow [19].

Let  $\psi_1 : I \rightarrow U_1$  be the first entry map for  $f$  (the same as for  $\hat{f}$ ) and  $\tau : I \rightarrow \mathbb{N} \cup \{0\}$ ,

$$\tau(x) = \inf\{k \geq 0 : f^k(x) \in U_1\},$$

so that  $\psi_1(x) = \hat{f}^{\tau(x)}(x) = f^{\tau(x)}(x)$ .

It follows from Lemma 3.15 that  $\int_I \tau \, dm \lesssim 1$ . Since  $f(\partial I) \subset \partial I$  and  $f^j(\partial U_1) \cap U_1 = \emptyset$  for all  $j$ , every branch of  $\psi_1$  is mapped diffeomorphically on  $U_1$ . By Lemma 3.9,  $\psi_1$  has universally bounded distortion.

Write  $m = \sum_{J \in B} m(J)m_J$ , where  $B$  is the set of all branches of  $\psi_1$  and  $m_J$  is the normalized to probability restriction of  $m$  to  $J$ . For each  $J$ , the probability measure  $f_*^{\tau(J)}m_J$  is supported on  $U_1$ , and due to the bounded distortion, it is regular in the sense of [19], with the regularity constant ( $R'$  in [19]) independent of  $t$ . Thus  $m$  is forward regular. The jump function  $\tau : B \rightarrow \mathbb{N} \cup \{0\}$  has bounded (uniformly in  $t$ ) first moment:  $\sum_{J \in B} m(J)\tau(J) \lesssim 1$ .

Let  $X_n$  and  $Y_n$  the discrete time random processes given by  $\sum_{j=0}^{n-1} \varphi \circ \hat{f}^j$  on the probability spaces  $(I, m)$  and  $(I, \hat{\mu})$  respectively. By [19, Thm. 2.5], there is a coupling of  $X_n$  and  $Y_n$ , that is, there exists a probability space  $\Omega$  supporting random processes  $\{X'_n\}$  and  $\{Y'_n\}$ , equal in distribution to  $\{X_n\}$  and  $\{Y_n\}$  respectively, such that

$$\mathbb{E} \left( \sup_{n \geq 0} |X'_n - Y'_n| \right) \lesssim \sup_I |\varphi|. \tag{6.5}$$

Bound (6.5), together with (6.4), implies our result.  $\square$

Let  $I_r = \cup\{J \in \mathcal{P} : J \text{ is red}\}$ .

**Lemma 6.7.** *There is a constant  $C > 0$ , independent of  $t$ , such that for all  $n \geq 0$ ,*

$$m\{x \in I : f^j(x) \notin I_r \text{ for all } j \leq n\} \gtrsim (1 - Ct)^n.$$

In particular, if  $n(t) = o(t^{-1})$ , then

$$\lim_{t \rightarrow 0} m\{x \in I : f^j(x) = \hat{f}^j(x) \text{ for all } j \leq n(t)\} = 1.$$

**Proof.** Let  $\tau : I \rightarrow \mathbb{N}$ ,

$$\tau(x) = \inf\{k \geq 1 : f^k(x) \in U_1\}$$

and  $g : I \rightarrow U_1, g(x) = f^{\tau(x)}(x)$ .

Observe that

$$m\{x \in I : f^j(x) \notin I_r \text{ for all } j \leq n\} \geq m\{x \in I : g^j(x) \notin I_r \text{ for all } j \leq n\}.$$

By Proposition 6.2, all branches of the map  $g$  in  $U_1 \setminus I_r$  are mapped diffeomorphically and with uniformly bounded distortion onto  $U_1$ . So are the branches in  $I \setminus U_1$ , following the argument for the first entry map  $\psi_1$  in the proof of Lemma 6.6. Proposition 6.2 guarantees that  $m(I_r) \lesssim t$ . Therefore,

$$m\{x \in I : g^n(x) \in I_r \mid g^j(x) \notin I_r \text{ for all } j < n\} \lesssim \frac{m(I_r)}{m(U_1)} \lesssim t.$$

The result follows.  $\square$

**Lemma 6.8.** *For all Lipschitz  $\varphi : I \rightarrow \mathbb{R}$ , we have  $\int \varphi \, d\hat{\mu}_t \rightarrow \int \varphi \, d\mu_0$  as  $t \rightarrow 0$ .*

**Proof.** For every (fixed)  $n \geq 1$ , the map  $(x, t) \mapsto f_t^n(x)$  is continuous. Thus

$$\sup_I \left| \frac{1}{n} \sum_{j=0}^{n-1} \varphi \circ f_t^j - \frac{1}{n} \sum_{j=0}^{n-1} \varphi \circ f_0^j \right| \rightarrow 0 \quad \text{as } t \rightarrow 0.$$

By Lemma 6.7, as  $t \rightarrow 0$ ,

$$\begin{aligned} \int_I \left| \frac{1}{n} \sum_{j=0}^{n-1} \varphi \circ \hat{f}_t^j - \frac{1}{n} \sum_{j=0}^{n-1} \varphi \circ f_0^j \right| dm &\leq \sup_I \left| \frac{1}{n} \sum_{j=0}^{n-1} \varphi \circ f_t^j - \frac{1}{n} \sum_{j=0}^{n-1} \varphi \circ f_0^j \right| \\ &+ 2 \sup_I |\varphi| m\{x \in I : f_t^j(x) = \hat{f}_t^j(x) \text{ for all } j \leq n\} = o(1). \end{aligned}$$

By Lemma 6.6,

$$\begin{aligned} & \left| \int \varphi d\hat{\mu}_0 - \int \varphi d\hat{\mu}_t \right| \\ & \lesssim \int_I \left| \frac{1}{n} \sum_{j=0}^{n-1} \varphi \circ \hat{f}_t^j - \frac{1}{n} \sum_{j=0}^{n-1} \varphi \circ f_0^j \right| dm + n^{-1/2} |\varphi|_{\text{Lip}} \\ & = o(1) + n^{-1/2} |\varphi|_{\text{Lip}}. \end{aligned}$$

Since we can fix  $n$  arbitrarily large, the result follows.  $\square$

### Declaration of competing interest

No competing interest.

### Appendix A. Moment estimates for nonuniformly expanding maps

Let  $(M, d)$  be a bounded metric space with a map  $f: M \rightarrow M$ . Suppose that  $Y \subset M$  and  $m$  is a Borel probability measure on  $Y$ . Suppose that  $\alpha$  is a finite or countable partition of  $Y$  (up to a zero measure set) with  $m(a) > 0$  for all  $a \in \alpha$ . We require that there exist an integrable function  $\tau: Y \rightarrow \{1, 2, \dots\}$ , constant on each  $a \in \alpha$  with value  $\tau(a)$ , and constants  $\lambda > 1$ ,  $K > 0$  and  $\eta \in (0, 1]$  such that for each  $a \in \alpha$ ,

- $F = f^\tau$  restricts to a (measure-theoretic) bijection from  $a$  to  $Y$ ;
- $d(F(x), F(y)) \geq \lambda d(x, y)$  for all  $x, y \in a$ ;
- $d(f^\ell(x), f^\ell(y)) \leq K d(F(x), F(y))$  for all  $x, y \in a$  and  $0 \leq \ell \leq \tau(a)$ ;
- the inverse Jacobian  $\zeta = \frac{dm}{dm \circ F}$  of the restriction  $F: a \rightarrow Y$  satisfies

$$|\log \zeta(x) - \log \zeta(y)| \leq K d(F(x), F(y))^\eta$$

for all  $x, y \in a$ .

We say that  $f: M \rightarrow M$  as above is a *nonuniformly expanding map*. We refer to  $Y$  as the *inducing set*, to  $\tau$  as the *inducing time* and to  $F$  as the *induced map*.

We assume that  $\int_Y \tau^2 dm < \infty$ . We use  $C$  to denote various positive constants which depend continuously (only) on  $\eta, K, \lambda, \text{diam } M$  and  $\int_Y \tau^2 dm$ .

**Lemma A.1** ([20, Prop. 2.5]). *There exists a unique  $F$ -invariant probability measure  $\mu_Y$  on  $Y$ , absolutely continuous with respect to  $m$ , and*

$$C^{-1} \leq \frac{d\mu_Y}{dm} \leq C.$$

Define a Young tower

$$\Delta = \{(y, \ell) \in Y \times \mathbb{Z} : 0 \leq \ell < \tau(y)\}$$

with a tower map  $T: \Delta \rightarrow \Delta$ ,

$$T(y, \ell) = \begin{cases} (y, \ell + 1), & \ell < \tau(y) - 1, \\ (F(y), 0), & \ell = \tau(y) - 1, \end{cases}$$

and a projection  $\pi: \Delta \rightarrow M$ ,  $\pi(y, \ell) = f^\ell(y)$ . Then  $\pi$  is a semi-conjugacy between  $T: \Delta \rightarrow \Delta$  and  $f: M \rightarrow M$ , i.e.  $\pi \circ T = f \circ \pi$ .

The measure

$$\mu_\Delta = \frac{\mu_Y \times \text{counting}}{\int \tau d\mu_Y}$$

is a  $T$ -invariant probability measure on  $\Delta$ , and  $\mu = \pi_*\mu_\Delta$  is an  $f$ -invariant probability measure on  $M$ .

Suppose that  $\varphi: M \rightarrow \mathbb{R}$ . Define

$$|\varphi|_\eta = \sup_{x \neq y \in M} \frac{|\varphi(y) - \varphi(x)|}{d(x, y)^\eta}, \quad |\varphi|_\infty = \sup_{x \in M} |\varphi(x)|, \quad \|\varphi\|_\eta = |\varphi|_\eta + |\varphi|_\infty. \tag{A.1}$$

We define similarly  $|\cdot|_\eta, |\cdot|_\infty$  and  $\|\cdot\|_\eta$  for functions  $\varphi: Y \rightarrow \mathbb{R}$ .

**Lemma A.2.** *Let  $\bar{\tau} = \int_Y \tau d\mu_Y$  and  $\tau_k = \sum_{j=0}^{k-1} \tau \circ F$ . Then*

$$|\tau_k - k\bar{\tau}|_{L^2(\mu_Y)} \leq Ck^{-1/2}.$$

**Proof.** Let  $P: L^1(\mu_Y) \rightarrow L^1(\mu_Y)$  denote the transfer operator corresponding to  $F$  and  $\mu_Y$ , so  $\int_Y v \circ F w d\mu_Y = \int_Y v Pw d\mu_Y$  for all  $v \in L^\infty$  and  $w \in L^1$ .

Let  $\varphi = \tau - \bar{\tau}$ . It is a direct verification that  $\|P\varphi\|_\eta \leq C$ . Thus, by [20, Cor. 2.4],  $\|P^k\varphi\|_\eta \leq C\gamma^k$  for all  $k \geq 1$ , where  $\gamma \in (0, 1)$  depends only on  $\lambda, K, \eta$  and  $\text{diam } M$ .

Finally,

$$\int_Y \left( \sum_{j=0}^{k-1} \varphi \circ F^j \right)^2 d\mu_Y \leq k \int_Y \varphi^2 d\mu_Y + 2k \sum_{j=1}^\infty \left| \int_Y \varphi \circ F^j \varphi d\mu_Y \right| \leq Ck.$$

The result follows.  $\square$

**Lemma A.3** ([22, Cor. 2.10]). *For all  $\varphi: M \rightarrow \mathbb{R}$  and  $n \geq 0$ ,*

$$\left| \sup_{k \leq n} \left| \sum_{j=0}^{k-1} \varphi \circ f^j - k \int \varphi d\mu \right| \right|_{L^2(\mu)} \leq C\|\varphi\|_\eta n^{1/2}.$$

Observe that  $\frac{dm}{d\mu} \leq C$ . Thus

**Corollary A.4.** *For all  $\varphi: M \rightarrow \mathbb{R}$  and  $n \geq 0$ ,*

$$\left| \sup_{k \leq n} \left| \sum_{j=0}^{k-1} \varphi \circ f^j - k \int \varphi d\mu \right| \right|_{L^2(m)} \leq C\|\varphi\|_\eta n^{1/2}.$$

We define a metric  $d_\Delta$  on  $\Delta$  by

$$d_\Delta((y, \ell), (y', \ell')) = \begin{cases} d(y, y'), & \ell = \ell'; \\ \text{diam } M, & \text{otherwise.} \end{cases}$$

Define  $|\cdot|_\eta, |\cdot|_\infty$  and  $\|\cdot\|_\eta$  for functions on  $\Delta$  similarly to (A.1).

**Remark A.5.**  $T: \Delta \rightarrow \Delta$  is itself a nonuniformly expanding map. Thus for all  $\psi: \Delta \rightarrow \mathbb{R}$ ,

$$\left| \sup_{k \leq n} \left| \sum_{j=0}^{k-1} \psi \circ T^j - k \int \psi d\mu_\Delta \right| \right|_{L^2(\mu_\Delta)} \leq C\|\psi\|_\eta n^{1/2}.$$

**References**

[1] J.F. Alves, H. Vilarinho, Strong stochastic stability for non-uniformly expanding maps, *Ergod. Theory Dyn. Syst.* 33 (3) (2013) 647–692.  
 [2] J.F. Alves, M. Carvalho, J.M. Freitas, Statistical stability for Hénon maps of the Benedicks-Carleson type, *Ann. Inst. Henri Poincaré, Anal. Non Linéaire* 27 (2) (2010) 595–637.  
 [3] J.F. Alves, M. Viana, Statistical stability for robust classes of maps with non-uniform expansion, *Ergod. Theory Dyn. Syst.* 22 (1) (2002) 1–32.

- [4] A.A. Andronov, L. Pontrjagin, *Systems grossiers*, Dokl. Akad. Nauk. USSR 14 (1937) 247–251.
- [5] V. Baladi, M. Benedicks, V. Maume-Deschamps, Almost sure rates of mixing for i.i.d. unimodal maps, *Ann. Sci. Éc. Norm. Supér.* (4) 35 (1) (2002) 77–126.
- [6] V. Baladi, M. Benedicks, D. Schnellmann, Whitney-Hölder continuity of the SRB measure for transversal families of smooth unimodal maps, *Invent. Math.* 201 (3) (2015) 773–844.
- [7] V. Baladi, M. Viana, Strong stochastic stability and rate of mixing for unimodal maps, *Ann. Sci. Éc. Norm. Supér.* (4) 29 (4) (1996) 483–517.
- [8] M. Benedicks, L. Carleson, The dynamics of the Hénon map, *Ann. Math.* (2) 133 (1) (1991) 73–169.
- [9] M. Benedicks, L.-S. Young, Absolutely continuous invariant measures and random perturbations for certain one-dimensional maps, *Ergod. Theory Dyn. Syst.* 12 (1) (1992) 13–37.
- [10] W. de Melo, S. van Strien, *One-Dimensional Dynamics*, *Ergebnisse der Mathematik und ihrer Grenzgebiete* (3) (Results in Mathematics and Related Areas (3)), vol. 25, Springer-Verlag, Berlin, 1993.
- [11] N. Dobbs, M. Todd, Free energy jumps up, preprint, arXiv:1512.09245, 2015.
- [12] D. Dolgopyat, On dynamics of mostly contracting diffeomorphisms, *Commun. Math. Phys.* 213 (1) (2000) 181–201.
- [13] J.M. Freitas, M. Todd, The statistical stability of equilibrium states for interval maps, *Nonlinearity* 22 (2) (2009) 259–281.
- [14] J. Graczyk, D. Sands, G. Świątek, Metric attractors for smooth unimodal maps, *Ann. Math.* (2) 159 (2) (2004) 725–740.
- [15] G.A. Gottwald, J.P. Wormell, J. Wouters, On spurious detection of linear response and misuse of the fluctuation-dissipation theorem in finite time series, *Physica D* 331 (2016) 89–101.
- [16] F. Hofbauer, G. Keller, Quadratic maps without asymptotic measure, *Commun. Math. Phys.* 127 (2) (1990) 319–337.
- [17] M.V. Jakobson, Absolutely continuous invariant measures for one-parameter families of one-dimensional maps, *Commun. Math. Phys.* 81 (1) (1981) 39–88.
- [18] G. Keller, Stochastic stability in some chaotic dynamical systems, *Monatshefte Math.* 94 (4) (1982) 313–333.
- [19] A. Korepanov, Equidistribution for nonuniformly expanding dynamical systems, and application to the almost sure invariance principle, *Commun. Math. Phys.* 359 (3) (2018) 1123–1138.
- [20] A. Korepanov, Z. Kosloff, I. Melbourne, Explicit coupling argument for nonuniformly hyperbolic transformations, *Proc. Edinb. Math. Soc.* 149 (1) (2019) 101–130.
- [21] A. Korepanov, Z. Kosloff, I. Melbourne, Averaging and rates of averaging for uniform families of deterministic fast-slow skew product systems, *Stud. Math.* 238 (1) (2017) 59–89.
- [22] A. Korepanov, Z. Kosloff, I. Melbourne, Martingale-coboundary decomposition for families of dynamical systems, *Ann. Inst. Henri Poincaré, Anal. Non Linéaire* 35 (4) (2018) 859–885.
- [23] G. Levin, On an analytic approach to the Fatou conjecture, *Fundam. Math.* 171 (2) (2002) 177–196.
- [24] S. Li, Q. Wang, The slow recurrence and stochastic stability of unimodal interval maps with wild attractors, *Nonlinearity* 26 (6) (2013) 1623–1637.
- [25] M. Lyubich, Almost every real quadratic map is either regular or stochastic, *Ann. Math.* (2) 156 (1) (2002) 1–78.
- [26] R. Mañé, Hyperbolicity, sinks and measure in one dimensional dynamics, *Commun. Math. Phys.* 100 (1985) 495–524.
- [27] M. Martens, Distortion results and invariant Cantor sets of unimodal maps, *Ergod. Theory Dyn. Syst.* 14 (2) (1994) 331–349.
- [28] M. Misiurewicz, Absolutely continuous measures for certain maps of an interval, *Publ. Math. Inst. Hautes Études Sci.* 53 (1981) 17–51.
- [29] T. Nowicki, D. Sands, Non-uniform hyperbolicity and universal bounds for  $S$ -unimodal maps, *Invent. Math.* 132 (3) (1998) 633–680.
- [30] M.M. Peixoto, Structural stability on two-dimensional manifolds, *Topology* 1 (1962) 101–120.
- [31] F. Przytycki, J. Rivera-Letelier, Geometric pressure for multimodal maps of the interval, *Mem. Am. Math. Soc.* 259 (1246) (2019).
- [32] D. Ruelle, Differentiating the absolutely continuous invariant measure of an interval map  $f$  with respect to  $f$ , *Commun. Math. Phys.* 258 (2) (2005) 445–453.
- [33] D. Ruelle, Structure and  $f$ -dependence of the A.C.I.M. for a unimodal map  $f$  is Misiurewicz type, *Commun. Math. Phys.* 287 (3) (2009) 1039–1070.
- [34] D. Sands, Misiurewicz maps are rare, *Commun. Math. Phys.* 197 (1) (1998) 109–129.
- [35] W. Shen, On stochastic stability of non-uniformly expanding interval maps, *Proc. Lond. Math. Soc.* (3) 107 (5) (2013) 1091–1134.
- [36] W. Shen, S. van Strien, On stochastic stability of expanding circle maps with neutral fixed points, *Dyn. Syst.* 28 (3) (2013) 423–452.
- [37] S. Smale, What is global analysis?, *Am. Math. Mon.* 76 (1969) 4–9.
- [38] H. Thunberg, Unfolding of chaotic unimodal maps and the parameter dependence of natural measures, *Nonlinearity* 14 (2) (2001) 323–337.
- [39] M. Tsujii, Positive Lyapunov exponents in families of one-dimensional dynamical systems, *Invent. Math.* 111 (1) (1993) 113–137.
- [40] M. Tsujii, On continuity of Bowen-Ruelle-Sinai measures in families of one-dimensional maps, *Commun. Math. Phys.* 177 (1) (1996) 1–11.