

FREE ENERGY JUMPS UP

NEIL DOBBS AND MIKE TODD

ABSTRACT. We study continuity, and lack thereof, of thermodynamical properties for one-dimensional dynamical systems. Under quite general hypotheses, the free energy is shown to be almost upper-semicontinuous: some normalised component of a limit measure will have free energy at least that of the limit of the free energies. From this, we deduce results concerning existence and continuity of equilibrium states (including statistical stability). Metric entropy, not semi-continuous as a general multimodal map varies, is shown to be upper semi-continuous under an appropriate hypothesis on critical orbits. Equilibrium states vary continuously, under mild hypotheses, as one varies the parameter and the map. We give a general method for constructing induced maps which automatically give strong exponential tail estimates. This allows us to recover, and further generalise, recent results concerning statistical properties (decay of correlations, etc.). Counterexamples to statistical stability are given which also show sharpness of the main results.

CONTENTS

1. Introduction	2
2. Method and structure	17
3. Piecewise preliminaries	20
4. Combinatorics in the Hofbauer extension	22
5. Coding convergence	28
6. Extension embedding	29
7. Dynamics of rare returns	30
8. Lifting and spreading measures and Abramov's formula	32
9. Light limit measures and entropy	35
10. Non-positive Schwarzian derivative	38
11. Almost upper-semicontinuity of the free energy	41
12. Katok theory	46
13. Ancillary proofs	48

Date: March 6, 2018.

MT was partially supported by FCT grant SFRH/BPD/26521/2006 and NSF grants DMS 0606343 and DMS 0908093. ND was supported by ERC Bridges project, the Academy of Finland CoE in Analysis and Dynamics Research and an IBM Goldstine fellowship.

14. Thermodynamics for induced maps	50
15. Instability for Collet-Eckmann maps	58
16. Positive entropy does not imply statistical stability	66
References	78

1. INTRODUCTION

An ideal gas is a collection of particles interacting solely via random collisions. Keeping track of individual particles is neither important nor feasible. Rather, macroscopic *thermodynamic* quantities (such as pressure, temperature, free energy) describe the system. One can use statistical mechanics to study dependence of equilibrium states and of thermodynamic quantities on parameters. In a deterministic (and for our purposes, discrete-time) dynamical system $f : X \rightarrow X$, orbits of nearby points remain close for a certain amount of time. If the system is expanding, the orbits diverge and become decorrelated. If this decorrelation occurs sufficiently quickly, the system exhibits random characteristics and chaotic behaviour. Individual orbits lose importance, and macroscopic, typical, statistical properties are of interest. In the 1970s, Sinai, Ruelle and Bowen [Si, Ru1, Bo] introduced *thermodynamic formalism* to dynamical systems to remarkable success¹.

In dynamical systems, questions of existence of equilibrium states and continuous dependence of the equilibrium state on parameters (aka *statistical stability*) are of prime importance. Uniformly hyperbolic systems are well-understood: existence holds and one even has smooth dependence of the equilibrium state on parameters [KKPW, Co, Ru3]. Non-hyperbolic systems are harder to study. Existence need not hold; even when it does, dependence need not be continuous, as we shall see. This paper contributes a new tool to the study of such questions in the context of one-dimensional dynamics. We give natural and general proofs of new results concerning statistical stability and linear response (topics studied in [K1, RS, Ts2, F, Ar, Ba1, FT, Ru5, BaS2, BBS, AS, VV]) or lack thereof, and we generalise recent work concerning existence and uniqueness of equilibrium states [IT1, PR1].

An equilibrium state is, roughly speaking, an invariant probability measure maximising the free energy. The naive approach to find an equilibrium state—taking a sequence of measures whose free energy converges to the supremum—does not work in general. Failure occurs because the free energy does not depend semicontinuously on the measure. We show that this unsophisticated approach has its merits, thanks to some hidden semicontinuity properties.

Our techniques lead us to propose (and prove) the following principle. For a large class of maps and potentials, *the free energy is almost upper-semicontinuous*. Let us explain briefly. Let f be a piecewise-smooth map of the unit interval I say, let μ be an ergodic invariant probability measure with entropy $h(\mu)$, and let $\varphi : I \rightarrow \overline{\mathbb{R}}$ be a potential function. The free energy $E(\mu)$ of μ is defined as $h(\mu) + \int \varphi d\mu$. Let E_+ denote the supremum of $E(\mu)$ (or equivalently of $\int \varphi d\mu$) over ergodic invariant probability measures supported on repelling periodic orbits. Take a convergent sequence of ergodic invariant probability measures μ_n with uniformly positive

¹A striking early result of Ruelle concerned Hausdorff dimension of Julia sets of quadratic complex polynomials [Ru2], a subject with, *a priori*, no connection to physics.

entropies and with $\liminf_n E(\mu_n) \geq E_+$. Then some component of the limit measure, when normalised, should have free energy at least $\limsup_n E(\mu_n)$.

This powerful principle is somewhat surprising, as (for potentials unbounded above, like $-t \log |Df|$, $t > 0$, when f has a critical point), $\mu \mapsto \int \varphi d\mu$ is often lower-semicontinuous, while (for a fixed map) entropy is upper-semicontinuous. *A priori* and in fact, the free energy is neither upper- nor lower- semicontinuous, yet we have some sort of hidden upper-semicontinuity.

What may happen is the following. Lower-semicontinuity of the Lyapunov exponent occurs when diminishing mass approaches the critical set and disappears in the limit. A typical orbit passing near the critical set leads to a bound period (not quite that of [BC]) where there is no freedom in the subsequent stretch of orbit, until one reaches the large scale again. Such orbital stretches considered as a whole make very little contribution to the free energy, so losing them in the limit and normalising the remainder (which requires positive entropy) does not cause free energy to drop. One may hope that some form of almost upper-semicontinuity will hold in higher dimensions; currently we do not have the techniques at our disposition to show this.

Hidden semicontinuity, which extends to families of maps, allows us to give strikingly straightforward proofs of existence and (almost) continuous dependence of equilibrium states, generalising many previous works. As a bonus, our techniques often give important statistical properties (in particular, the *almost sure invariance principle* and *exponential decay of correlations*) and we derive regularity of pressure functions.

For sequences of maps, even metric entropy (of invariant measures) is not necessarily semicontinuous [M2]. This would appear to render our endeavour hopeless. However, our hypothesis, *decreasing critical relations*, guarantees upper-semicontinuity of entropy. It is relatively weak: for the quadratic family, only super-attracting parameters have an extra critical relation in the limit, for example.

Changing tack, a lower bound on the entropy on its own is not enough (contrary to prior suspicions) to guarantee continuity of equilibrium states. We find sequences of non-renormalisable, post-critically finite maps with absolutely continuous invariant probability measures (*acips*) with entropy bounded away from zero, for which the limit measure is a convex combination of the acip for the limit map and a Dirac mass on a repelling fixed point.

Worse things happen without the lower bound on the entropy. Motivated by Ruelle's paper [Ru5] showing smoothness of dependence of the acip on parameter for a family of Misiurewicz maps which remain *in the same topological class*, we consider what happens if ever one leaves the class: Even restricting to non-renormalisable Misiurewicz parameters, the limit measure can be more or less anything, see Theorem 1.30.

1.1. Equilibrium states. Given a dynamical system $f : I \rightarrow I$ on a topological space I , and a *potential* $\varphi : I \rightarrow [-\infty, \infty]$, both of which preserve the Borel structure, we define the (standard *variational*) *pressure* of (I, f, φ) to be

$$\bar{P}_f(\varphi) := \sup \left\{ h(\mu) + \int \varphi d\mu : \mu \in \bar{\mathcal{M}} \text{ and } \int \varphi d\mu > -\infty \right\},$$

where $h(\mu) = h_f(\mu)$ denotes the (metric) entropy of μ and $\bar{\mathcal{M}} = \bar{\mathcal{M}}_f$ denotes the set of ergodic, f -invariant, probability measures. Spaces of measures are equipped with the weak* topology.

Henceforth, let I be a bounded, non-degenerate interval. For a piecewise-smooth one-dimensional map f and $\mu \in \overline{\mathcal{M}}_f$, the Lyapunov exponent $\lambda(\mu)$ of μ is defined as $\int \log |Df| d\mu$. Ergodic measures with negative Lyapunov exponent have zero entropy, are typically finite in number and are supported on attracting periodic orbits (at least in the smooth case, see [Ri, Proposition A.1]). They do not contribute much by way of knowledge. Often they are defined away by, for example, assuming that the system is transitive. We shall rather consider the following pressure function,

$$P(\varphi) = P_f(\varphi) := \sup_{\mu \in \mathcal{M}_f} \left\{ h(\mu) + \int \varphi d\mu : \int \varphi d\mu > -\infty \right\},$$

where

$$\mathcal{M} = \mathcal{M}_f := \left\{ \mu \in \overline{\mathcal{M}}_f : \int \log |Df| d\mu \geq 0 \right\}.$$

Results for \overline{P} can be deduced from those for P . As is traditional in dynamical systems [K3], for an invariant measure, set

$$E(\mu) = E(f, \mu, \varphi) := h_f(\mu) + \int \varphi d\mu$$

and call it the *free energy* of μ (with respect to (I, f, φ)). Note $E(\mu + \alpha\mu') = E(\mu) + \alpha E(\mu')$ for $\alpha \in \mathbb{R}^+$. We define

$$E_+ = E_+(f) = E_+(f, \varphi) := \sup \left\{ E(\mu) : \begin{array}{l} \mu \in \mathcal{M}_f \text{ is supported on a} \\ \text{repelling periodic orbit} \end{array} \right\}.$$

If a measure $\mu \in \mathcal{M}$ *maximises* the free energy, thus satisfying $h(\mu) + \int \varphi d\mu = P(\varphi)$, then we call μ an *equilibrium state* for (I, f, φ) . Note we require equilibrium states to be ergodic.

The family of potentials $\{-t \log |Df| : t \in \mathbb{R}\}$ is of especial interest; the integral $\int \log |Df| d\mu$ is the Lyapunov exponent of the measure. For example, since an f -invariant measure which is absolutely continuous with respect to Lebesgue measure describes the statistical behaviour of a set of points of positive Lebesgue measure, such measures are of great import. It was shown in [L] that if f is C^2 with non-flat critical points, then a measure μ with positive entropy is an acip if and only if it is an equilibrium state for the potential $-\log |Df|$ (the non-flat condition is unnecessary [D2]). For quadratic maps, Lebesgue measure is ergodic and any acip has positive entropy [BL]. Moreover, equilibrium states of other potentials $-t \log |Df|$ give information on the Lyapunov spectrum, see for example [IT2].

The function $t \mapsto P(-t \log |Df|)$ is convex (it can be viewed as a supremum of lines $t \mapsto E(f, \mu, -t \log |Df|)$) and mostly smooth. A point t_0 where it is not analytic is called a *phase transition*. If the function is smooth at t and has an equilibrium state μ_t , then the slope of the pressure function is minus the Lyapunov exponent of μ_t . If $h(\mu_t) > 0$, the measure μ_t has maximal dimension (equal to $h(\mu_t)/\lambda(\mu_t)$) by the Dynamic Volume Lemma [HR]) among measures with the same Lyapunov exponent.

Let us set $P^0(\varphi) = P^0(f, \varphi) := \sup\{E(f, \mu, \varphi) : h(\mu) = 0 \text{ and } \mu \in \mathcal{M}_f\}$. Of course, $P^0(f, \varphi) \geq E_+(f, \varphi)$, the supremum in E_+ being taken over repelling orbits; for reasonable classes of maps one may expect equality. Consider

$$t^- := \inf \{t : P(-t \log |Df|) > P^0(-t \log |Df|)\},$$

$$t^+ := \sup \{t : P(-t \log |Df|) > P^0(-t \log |Df|)\}.$$

These need not be finite; t^+ especially depends strongly (often discontinuously) on the map f . There are interesting results on maps for which $t^+ \in (1, \infty)$ by Coronel and Rivera-Letelier

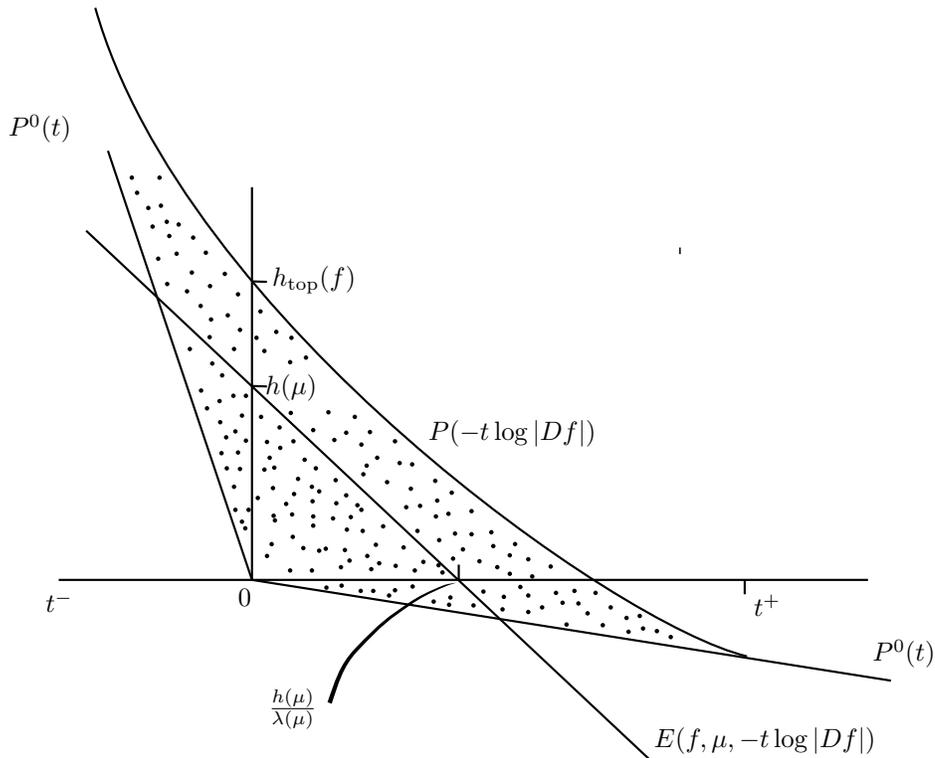


FIGURE 1. The zone of almost upper-semicontinuity of the free energy

[CR]. If we assume that f has positive topological entropy, then $P(0) > P^0(0)$, so $t^- < 0$ and $t^+ > 0$. For $t \in (t^-, t^+)$, $P(-t \log |Df|) > P^0(-t \log |Df|)$ and any equilibrium measure necessarily has positive entropy. Moreover, if $E(-t \log |Df|, \mu) > P^0(-t \log |Df|)$ then μ has positive entropy. One can think of the closure of the zone $P^0 < E < P$ as the region of possible application of our results, see Figure 1.

Existence of equilibrium states is non-trivial. For smooth interval maps with critical points, for example, the entropy depends upper-semicontinuously on the measure, but $\mu \mapsto -\int \log |Df| d\mu$ is lower-semicontinuous, so if one naively takes a convergent sequence of measures whose free energies converge to the pressure, the limit measure may not maximise the free energy. Limit measures need not be ergodic.

Definition 1.1. *Given a sequence of f -invariant probability measures, we call an f -invariant measure μ a light limit measure provided some subsequence converges to an f -invariant measure μ' , $\mu \ll \mu'$ and $\mu(I) = 1$.*

In the above, if μ' were ergodic, then μ would be μ' ; if not, then one could decompose μ' into two non-trivial f -invariant measures; each one, when normalised, would be a light limit measure. We do not require light limit measures to be ergodic.

There is a compatible extension of the definition of light limit measures to families of maps.

Definition 1.2. *Let $(f_k)_k$ be a convergent sequence of d -branched piecewise-monotone maps (see Definition 1.10) with d -branched limit f_0 and with ergodic f_k -invariant probability measures*

μ_k . We say an f_0 -invariant measure μ is a light limit measure provided some subsequence of the μ_k converges to an f_0 -invariant measure μ' , $\mu \ll \mu'$ and $\mu(I) = 1$.

1.2. Applications to the quadratic family. The quadratic (or logistic) family

$$\mathcal{F}_Q := \{f_a : x \mapsto ax(1-x), \quad a \in [1, 4]\}$$

is the primordial test-bed for ideas in non-linear, non-hyperbolic dynamics and is the subject of vast literature, see [MeS] for example. Phenomena from quadratic dynamics appear in the real world, for example period doubling [LLF]. Our results apply (and are new) in this context. We present a simplified version of one theorem, with the intention of conveying the concept and motivating the reader to delve further into the paper. We first require some definitions.

Given $\varepsilon > 0$ and a potential φ , let us write

$$\mathcal{M}_f^\varepsilon(\varphi) := \{\mu \in \mathcal{M}_f : E(f, \mu, \varphi) \geq E_+(\varphi) \quad \text{and} \quad h(\mu) \geq \varepsilon\}.$$

This is the class of measures from which one can hope for some good behaviour. Limits of ergodic measures need not be ergodic. Let us call an f -invariant measure μ_* *hyperbolic* if μ_* -almost every point x has positive (pointwise) Lyapunov exponent $\lim_{n \rightarrow \infty} \frac{1}{n} \log |Df^n(x)| > 0$.

Definition 1.3. *We say the free energy is almost upper-semicontinuous for the quadratic family \mathcal{F}_Q and the potentials $-t \log |Df|$ ($f \in \mathcal{F}_Q$, $t \in \mathbb{R}$) if, for every $\varepsilon > 0$, for every sequence $(a_k, t_k, \mu_k)_k$ such that*

- $(a_k)_k$ converges to $a_0 \in [1, 4]$ as $k \rightarrow \infty$,
- $(t_k)_k$ converges to $t_0 \in \mathbb{R}$ as $k \rightarrow \infty$
- and the measures $(\mu_k)_k$ satisfy

$$\mu_k \in \mathcal{M}_{f_{a_k}}^\varepsilon(-t_k \log |Df_{a_k}|),$$

there exists a hyperbolic light limit measure μ_* with

$$E(f_{a_0}, \mu_*, -t_0 \log |Df_{a_0}|) \geq \limsup_k E(f_{a_k}, \mu_k, -t_k \log |Df_{a_k}|).$$

Theorem 1.4. *The free energy is almost upper-semicontinuous for the quadratic family and the potentials $-t \log |Df|$.*

While we show much more later in the introduction, let us present a consequence (also true in much greater generality) to illustrate the power of this idea.

One can always take a sequence of measures with free energies converging to the pressure. Sometimes (in particular, if $t \in (t^-, t^+)$) all measures with large free energy have positive entropy. Applying the theorem for a fixed map f , one easily recovers the essence of the main theorem (A) of [IT1], for example:

Corollary 1.5. *For $f \in \mathcal{F}_Q$ and $t \in (-\infty, t^+)$, there exists an equilibrium state μ_t for the potential $-t \log |Df|$. For $t \in (t^-, t^+)$, there exists such μ_t with positive entropy.*

If statistical properties of a dynamical system are described by invariant measures, it is natural to ask how these measures (and properties) vary as the parameters of a dynamical system change. For example, if a quadratic map has a hyperbolic attracting periodic orbit (*i.e.*, with negative Lyapunov exponent), then all nearby maps do too, and the periodic orbit moves smoothly with the parameter. The equidistributions on the periodic attractors are ergodic invariant probability measures which describe the typical long-term dynamics of almost every

point, since the basins of attraction have full measure [MeS]. As these measures vary continuously, we say that statistical stability holds at parameters with hyperbolic attracting periodic orbits.

The question then arises as to the dependence of absolutely continuous invariant probability measures as a function of parameter. Let \mathcal{F} be a subset of \mathcal{F}_Q and suppose that for all $f \in \mathcal{F}$, f has an acip (acips are always unique for maps in the quadratic family). We will say that acips are *statistically stable* in \mathcal{F} if the acip depends continuously on the parameter. Rychlik and Sorets [RS] proved statistical stability of acips at Misiurewicz maps, a zero measure set of parameters. For a positive measure set $\mathcal{F} \subset \mathcal{F}_Q$, Tsujii used Benedicks-Carleson techniques to show statistical stability of acips in \mathcal{F} [Ts2]. This was extended by Freitas and Todd to the more general case where \mathcal{F} is any set of non-renormalisable parameters satisfying a uniform summability condition for the derivatives along the critical orbit [FT]. Recently, Baladi *et al* showed Hölder continuity in Tsujii's setting [BBS]. On the other hand, Tsujii and Thunberg [Ts2, Th] showed how renormalisation can lead to instability; we shall present further obstructions to statistical stability, but first we shall give a positive result.

Definition 1.6 (Statistical quasistability). *For each $\varepsilon > 0$, denote by $\mathcal{F}_Q^\varepsilon$ the subfamily of maps f in \mathcal{F}_Q for which f has an acip with entropy $\geq \varepsilon$. We shall say acips are statistically quasistable in the quadratic family \mathcal{F}_Q if, for each $\varepsilon > 0$, for any $f_0 \in \mathcal{F}_Q^\varepsilon$ and sequence $(f_k)_{k \geq 1}$ of maps in $\mathcal{F}_Q^\varepsilon$ converging to f_0 , some light limit measure of the corresponding sequence of acips is the acip for f_0 .*

Recalling that acips in the quadratic family are equilibrium states for the potential $-\log |Df|$, Theorem 1.4 implies the following result. Note that there is no assumption on the behaviour of the (derivatives along the) critical orbit.

Corollary 1.7. *Acips are statistically quasistable in the quadratic family \mathcal{F}_Q .*

There was hope in the community that statistical stability (and more) might hold for non-renormalisable Collet-Eckmann parameters (those for which the derivative grows exponentially along the critical orbit), see [Ba1, Conjecture B]. However, even for the better-behaved Misiurewicz parameters, one has instability [even worse, the limit measure can be almost anything at all, see Theorem 1.30]:

Theorem 1.8. *Statistical stability does not hold for acips for the subfamily of non-renormalisable Misiurewicz quadratic maps.*

It follows from the work of Tsujii [Ts1, Ts2], though justification is beyond the scope of this paragraph, that near (almost) every non-renormalisable Misiurewicz parameter there is a positive measure set of non-renormalisable Collet-Eckmann parameters whose acips are very close to that of the Misiurewicz parameter. Compare the discussion in §1.8.8. From this and Theorem 1.8, one can deduce the following.

Corollary 1.9. *Statistical stability does not hold for acips for any (relatively) full-measure subset of non-renormalisable Collet-Eckmann parameters in the quadratic family.*

Note that statistical quasistability is weaker than statistical stability, in that only some component of the (not necessarily ergodic) limit measure need be the acip of the limit map. While the examples of Theorem 1.30 have entropy decreasing to zero, uniform positive entropy will not imply stability, as we show in Theorem 1.31. One cannot expect better than quasistability

– for actual statistical stability one needs stronger hypotheses. Consider a sequence of (maps and) acips with uniform positive entropy. Theorem 1.4 says that some light limit measure will be an acip. However, another component of the limit measure may be an atom at a repelling fixed point for the limit map, for example.

1.3. Broader results and piecewise-monotone families. A quadratic map is a continuous 2-branched piecewise-monotone map. We shall study a much broader class of maps, admitting discontinuities. In this subsection, we present results concerning almost-semicontinuity and statistical stability for general piecewise-monotone families. There are also results pertaining to the thermodynamic formalism for an individual map.

Definition 1.10. For $d \geq 2$, a map f defined on a pairwise-disjoint collection I_1, I_2, \dots, I_d of non-degenerate subintervals of a bounded interval I is called a (d -branched) piecewise-monotone map if f maps each I_j into I and, on each I_j , f is continuous and strictly monotone.

We shall assume throughout the paper, without loss of generality, that $d \geq 2$. Otherwise the entropy of the map, and thus of any invariant measure, would be zero.

In the presence of more than one such map, we write $I_j(f)$ for I_j . A branch I_j of f is called *full* if it is mapped by f onto I . We do not specify here whether the subintervals I_j are open or closed. Since the domain of f need not be the whole interval I , iterates for some points may not be defined. However, for any invariant probability measure, almost every point has all iterates defined. Moreover, almost every point is recurrent.

Definition 1.11. We call \mathcal{F} a d -branched piecewise-monotone family on I if each $f \in \mathcal{F}$ is a d -branched piecewise-monotone map.

\mathcal{F} is called a piecewise-monotone family if it is a d -branched piecewise-monotone family for some $d \geq 2$.

Definition 1.12. Let \mathcal{F} be a d -branched piecewise-monotone family. We say a sequence of $f_k \in \mathcal{F}$ converges to f in \mathcal{F} as $k \rightarrow \infty$ if $f \in \mathcal{F}$ and there is a sequence of homeomorphisms $h_k : I \rightarrow I$ such that, for $1 \leq j \leq d$,

- $h_k(I_j(f_k)) = I_j(f)$,
- $h_k \circ f_k \circ h_k^{-1}$ converges in C^0 to f (on the domain of definition of f) as $k \rightarrow \infty$,
- h_k converges in C^0 to the identity on I as $k \rightarrow \infty$.

This definition is equivalent to uniform convergence on compact subsets of

$$\bigcup_j I'_j(f),$$

where $I'_j(f)$ denotes (momentarily) the interior of $I_j(f)$, provided the f_j and f are uniformly Lipschitz. For f as above, set

$$\mathcal{E}(f) := \bigcup_j \partial I_j.$$

The Hofbauer extension, a Markovian extension of an original system, was introduced in [H2]. This powerful idea lies behind a host of results in one-dimensional dynamics. The structure of the Hofbauer extension depends on the post-critical orbits. In recent work on statistical stability [FT] where Hofbauer extensions were used, the limit maps studied were assumed not to have critical orbits which intersect or self-intersect. That unsatisfactory assumption excludes

post-critically finite maps and many more; it was necessary to obtain convergence of Hofbauer extensions. We introduce the following definition to deal with this issue, eventually embedding Hofbauer extensions in some ambient space and examining convergence properties there.

Definition 1.13. *Let f be a piecewise-monotone map and let $c \in \mathcal{E}(f)$. Let $\sigma \in \{-1, +1\}$. Suppose there is some minimal $n \geq 1$ for which*

$$\lim_{\varepsilon \rightarrow 0^+} f^n(c + \sigma\varepsilon) = c' \in \mathcal{E}(f).$$

We call this a critical relation of order n . This is a purely topological condition, defined by (c, c', σ, n) . We say that a sequence (f_k) such that $f_k \rightarrow f_0$ as $k \rightarrow \infty$ has decreasing critical relations if to each critical relation (c_0, c'_0, σ, n) of f_0 there is a corresponding critical relation (c_k, c'_k, σ, n) of f_k , where $h_k^{-1}(c_0) = c_k$ and $h_k^{-1}(c'_0) = c'_k$.

Of course, a sequence will have decreasing critical relations if all maps from the sequence share the same critical relations. For multimodal maps, there may be many critical relations. This is not true for unimodal maps.

Remark 1.14 (Critical relations for unimodal maps). *Consider a quadratic map $f_a : x \mapsto ax(1-x)$. It maps the interval $[0, 1]$ into itself for $a \in (3, 4]$, the parameters of interest. Extend the domain of definition to $[-\varepsilon, 1 + \varepsilon]$ for some $\varepsilon > 0$. Then the boundary points are mapped outside the domain so they contribute no critical relations. The orbit of the critical point $\frac{1}{2}$ never leaves $[0, 1]$, so it can only cause a critical relation if it is periodic and (thus) super-attracting. These critical relations are negligible: If one is only interested in positive-entropy measures, no mass lies in the basin of attraction. Remove a small neighbourhood of $\frac{1}{2}$ contained in the basin of attraction and the critical relation disappears. If the neighbourhood shrinks to nothing quickly, as one changes the parameter, the conditions for being a piecewise-monotone family remain verified. We only used the C^1 nature of quadratic maps, so the same holds for general C^1 unimodal maps.*

As discussed in §1.8.7, some assumption regarding critical relations is necessary in our work. In particular, there are examples showing metric entropy is not necessarily upper-semicontinuous as one varies both map and measure. On the other hand, Raith has shown that for unimodal maps, the measure of maximal entropy varies continuously [Ra]. This admits a strong generalisation, proven in §9:

Theorem 1.15 (Upper-semicontinuity of entropy). *Let $(f_k)_{k \geq 0}$ be a sequence of d -branched maps converging to f_0 as $k \rightarrow \infty$ with decreasing critical relations. Suppose $(\mu_k)_{k \geq 1}$ is a convergent sequence of measures $\mu_k \in \mathcal{M}_{f_k}$. Then*

$$h \left(\lim_{k \rightarrow \infty} \mu_k \right) \geq \liminf_{k \rightarrow \infty} h(\mu_k).$$

Definition 1.16. *We say a piecewise-monotone map f has non-positive Schwarzian derivative if, on the interior of each I_j , f is a C^2 diffeomorphism and $1/\sqrt{|Df|}$ is convex. If the convexity is strict, the map has negative Schwarzian derivative.*

Non-positive [M1, p.24] or negative Schwarzian derivative [MeS, §IV.1] is a standard property which gives good distortion control for iterates via the Koebe Lemma (see Lemma 10.2). It is not especially restrictive ([Ko]). Maps from the quadratic family $f_a : x \mapsto ax(1-x)$ with $a \in (0, 4]$ are (2-branched) piecewise-monotone maps with negative Schwarzian derivative. One can readily check that maps with non-positive Schwarzian derivative have bounded derivative. The decreasing critical relations hypothesis allows one to obtain uniform Koebe space for some natural induced maps (see Lemma 5.4).

Definition 1.17. Let \mathcal{F}_{NSD} denote the class of piecewise-monotone families \mathcal{F} such that for each $f \in \mathcal{F} \in \mathcal{F}_{\text{NSD}}$, f has non-positive Schwarzian derivative and $\sup_{f \in \mathcal{F}} \sup |Df| < +\infty$.

By Definition 1.11, each element of \mathcal{F}_{NSD} is a d -branched family for some d . In our applications of this notion, we will often take some sequence $(f_k)_k \in \mathcal{F}_{\text{NSD}}$.

Theorem 1.18 (Free energy is almost upper-semicontinuous). *Let $(f_k)_{k \geq 0} \in \mathcal{F}_{\text{NSD}}$ be a sequence converging to f_0 as $k \rightarrow \infty$ with decreasing critical relations. Suppose that*

- (a) $t_k \rightarrow t_0$ as $k \rightarrow \infty$;
- (b) $(\mu_k)_{k \geq 1}$ is a convergent sequence of measures $\mu_k \in \mathcal{M}_{f_k}$;
- (c) $E(f_k, \mu_k, -t_k \log |Df_k|)$ converges to a limit E_L as $k \rightarrow \infty$;
- (d) $E_L \geq \liminf E_+(f_k, -t_k \log |Df_k|)$;
- (e) $\liminf_{k \rightarrow \infty} h(\mu_k) > 0$.

Then, writing $E(\mu_*)$ for $E(f_0, \mu_*, -t_0 \log |Df_0|)$, some light limit measure μ_* is hyperbolic and satisfies one of the following statements:

- $E(\mu_*) > E_L$;
- $\mu_* = \lim_{k \rightarrow \infty} \mu_k$, $\int \log |Df_0| d\mu_* = \lim_{k \rightarrow \infty} \int \log |Df_k| d\mu_k$,
 $h(\mu_*) \geq \liminf_{k \rightarrow \infty} h(\mu_k)$ and $E(\mu_*) \geq E_L$;
- $E(\mu_*) \geq E_L$ and $h(\mu_*) > \liminf_{k \rightarrow \infty} h(\mu_k)$.

If $E_L = P(-t_0 \log |Df_0|)$, one of the last two alternatives holds and some light limit measure is a positive entropy equilibrium state for this potential.

Theorem 1.18 is the principal technical result of the paper; its proof occupies §§3-11. Further results, detailed below and proven in §13, follow with little extra work.

Theorem 1.18 and Remark 1.14 imply the following corollary, equivalent to Theorem 1.4.

Corollary 1.19. *The statement of Theorem 1.18 holds for the quadratic family \mathcal{F}_Q , without the decreasing critical relations hypothesis.*

We shall later treat analyticity of the pressure function and statistical properties, but for now let us just consider existence of equilibrium states. Existence was shown in [IT1, Theorem A] and again in [PR1]. Compared with those settings, we allow discontinuities, parabolic points and holes ([PR1] allows holes), while both [IT1, PR1] impose a *non-flatness* condition on critical points (see also §1.8.4). The following corollary implies Corollary 1.5.

Corollary 1.20. *Given a d -branched map f with non-positive Schwarzian derivative, for $t \in (-\infty, t^+)$, there exists an equilibrium state μ_t for the potential $-t \log |Df|$. For $t \in (t^-, t^+)$, there exists such μ_t with positive entropy.*

From the pressure function, one can read off certain properties. For example, the following result generalises [IT1, Proposition 1.1].

Proposition 1.21. *The map $t \mapsto P(-t \log |Df|)$ is differentiable at t^+ if and only if there is no equilibrium state with positive entropy for the potential $-t^+ \log |Df|$. The same holds for t^- in place of t^+ .*

Concerning statistical stability of equilibrium states, we obtain strong, new results.

Theorem 1.22 (Smooth pressure and uniform pressure gap imply statistical stability). *Let $(f_k)_k \in \mathcal{F}_{\text{NSD}}$ be a sequence converging to f_0 as $k \rightarrow \infty$ with decreasing critical relations. Suppose*

- (a) $t_k \rightarrow t_0$ as $k \rightarrow \infty$;
- (b) for each $k \geq 0$, $P(f_k, -t_0 \log |Df_k|) \geq P^0(f_k, -t_0 \log |Df_k|) + \varepsilon$.

For $k \geq 0$, there is an equilibrium state μ_k for f_k and the potential $-t_k \log |Df_k|$. If $t \mapsto P(f_0, -t \log |Df_0|)$ is differentiable at t_0 , then any limit measure of the sequence $(\mu_k)_k$ is a convex combination of equilibrium measures for f_0 with the potential $-t_0 \log |Df_0|$.

Jumping ahead a little, for non-renormalisable quadratic maps, f is transitive on $J(f)$ so positive entropy equilibrium states are unique, see Theorem 1.27. Moreover the pressure is analytic on (t^-, t^+) , see Theorem 1.28. In the quadratic family \mathcal{F}_Q , we immediately obtain:

Corollary 1.23 (Statistical stability in the quadratic family). *Suppose that $f_0 \in \mathcal{F}_Q$ is non-renormalisable, $t_0 \in (t^-, t^+)$ and $P(f_0, -t_0 \log |Df|) > 0$. For (t, f) in a neighbourhood of (t_0, f_0) in $\mathbb{R} \times \mathcal{F}_Q$, an equilibrium state $\mu_{t,f}$ (for f and the potential $-t \log |Df|$) exists. The map $(t, f) \mapsto \mu_{t,f}$ is continuous at (t_0, f_0) .*

This significantly improves upon [FT], where stability was shown for t close to 1 under multiple further hypotheses, including non-uniform expansion.

1.4. $J(f)$ for a piecewise-monotone map f .

Definition 1.24. *Given a piecewise-monotone map $f : \cup_{j=1}^d I_j \rightarrow I$, the set of critical points is denoted*

$$\mathcal{E}(f) := \bigcup_j \partial I_j,$$

while the set of critical values is denoted

$$\mathcal{V}(f) := \bigcup_j \partial f(I_j).$$

Definition 1.25. *Given a piecewise-monotone map f , we denote by $J(f)$ the set of points x such that*

- x is recurrent, so $x \in \overline{\{f^n(x)\}_{n \geq 1}}$;
- x is accumulated on both sides by points from $\cup_{n \geq 0} f^{-n}(\mathcal{E}(f))$, that is, for all $\varepsilon > 0$, there exist $l, r \in \cup_{n \geq 0} f^{-n}(\mathcal{E}(f))$ with $x - \varepsilon < l < x < r < x + \varepsilon$;
- $f^n(x) \notin \mathcal{E}(f)$ and $f^{-n}(f^m(x)) \cap \mathcal{V}(f) = \emptyset$ for each $n, m \geq 0$;
- x is accumulated on both sides by points with the above three properties.

It is easy to check that $J(f)$ is forward-invariant, $f(J(f)) \subset J(f)$. For any ergodic, invariant, positive-entropy measure, the set $J(f)$ has full measure – we shall justify this statement in Lemma 3.1. On the other hand, some zero-entropy measures may give no mass to $J(f)$. This definition of $J(f)$ is related to, but more restrictive than, the definition of Julia set used in [Ri], the main differences being recurrence and that we require our accumulation from both sides. Of course, $J(f)$ is not a closed set.

Remark 1.26. Quadratic maps $f_a : x \mapsto ax(1-x)$ with $a \in (3, 4)$ are not transitive on the unit interval $[0, 1]$. However, for certain values of a , in particular when f_a is non-renormalisable, they may be transitive on the interval $[f_a^2(1/2), f_a(1/c)]$. In such cases, $\overline{J(f_a)}$ would coincide with this forward-invariant dynamical core $[f_a^2(1/2), f_a(1/2)]$ and f_a would be transitive on $J(f_a)$. The only ergodic invariant probability measure outside the core would be the Dirac mass at 0. If one extends the domain to $[-\varepsilon, 1 + \varepsilon]$, exclusion of isolated recurrent points means $0 \notin J(f_a)$.

1.5. Uniqueness of equilibrium states, analyticity of pressure, decay of correlations.

One could ask how many equilibrium states there are for a given potential, compare [Pi]. Depending on the potential and the map, the answer may be infinitely many (see the final paragraph of [D1]). Restricting to those with uniformly positive entropy, we shall provide a finite bound on the number and this bound only depends on the number of branches and the entropy. With a transitivity assumption (on the chaotic part of phase space), there is only one equilibrium state.

Theorem 1.27 (Bounds on the number of equilibrium states). *Given $d \geq 2, \varepsilon > 0$, there is a number $N \in \mathbb{N}$ such that the following holds. If f is a d -branched piecewise-monotone map with non-positive Schwarzian derivative and $t \in \mathbb{R}$, there are at most N equilibrium states with entropy greater than ε for the potential $-t \log |Df|$.*

If f is transitive on $J(f)$, then there is at most one equilibrium state with positive entropy for the potential $-t \log |Df|$.

Understanding the pressure function $t \mapsto P_f(t\varphi)$ and its smoothness properties can give information on the statistical properties (large deviations and multifractal spectra, for example) of the system I, f . We note that if there exists an equilibrium state μ_s for $s\varphi$, and $t \mapsto P(t\varphi)$ is differentiable at $t = s$ then

$$\left. \frac{\partial P(t\varphi)}{\partial t} \right|_{t=s} = \int \varphi d\mu_s. \quad (1.1)$$

The proofs of the following two results depend on thermodynamic formalism for shifts on an infinite alphabet. They correspond to [PR1, Theorem A]. Again recall, we do not exclude parabolic points and do not assume non-flatness of critical points.

Theorem 1.28. *Let f be a piecewise-monotone map with non-positive Schwarzian derivative. If f is transitive on $J(f)$, the pressure function $t \mapsto P(-t \log |Df|)$ is real-analytic on the interval (t^-, t^+) .*

Suppose that (I, f, μ) is an ergodic dynamical system. Following [MN, Theorem 1.2], we say that $\varphi : I \rightarrow \mathbb{R}$ satisfies the Almost Sure Invariance Principle (ASIP) if there exists $\gamma > 0$, a sequence of random variables $\{S_N\}_N$ and a Brownian motion W with variance $\sigma^2 \geq 0$ such that

$$\left\{ \sum_{j=0}^{N-1} \varphi \circ f^j \right\}_N = \{S_N\}_N \quad \text{in distribution}$$

and, almost everywhere,

$$S_N = W(N) + O(N^{\frac{1}{2}-\epsilon}) \quad \text{as } N \rightarrow \infty.$$

For a pair of function spaces $\mathcal{C}_1, \mathcal{C}_2$, we say that we have *decay of correlations*, against $(\mathcal{C}_1, \mathcal{C}_2)$ observables, if there exists a function $\rho : \mathbb{N} \rightarrow [0, \infty)$ such that $\rho(n) \rightarrow 0$ as $n \rightarrow \infty$ such that

for each pair $\varphi \in \mathcal{C}_1$, $\psi \in \mathcal{C}_2$ there exists a constant $C_{\varphi, \psi} > 0$ such that for any $n \in \mathbb{N}$,

$$\left| \int (\varphi \circ f^n) \psi \, d\mu - \int \varphi \, d\mu \int \psi \, d\mu \right| \leq C_{\varphi, \psi} \rho(n).$$

If $\rho(n) = O(e^{-\alpha n})$ for some $\alpha > 0$, we say that we have exponential decay of correlations.

For $\beta > 0$, let \mathcal{H}_β denote the set of Hölder continuous observables $\varphi : I \rightarrow \mathbb{R}$.

Theorem 1.29 (Decay of correlations and the ASIP). *Let $\beta > 0$ and let f be a piecewise-monotone map with non-positive Schwarzian derivative. Assume f is topologically mixing on $J(f)$. For each $t \in (t^-, t^+)$, the (unique) equilibrium state for the potential $-t \log |Df|$ has exponential decay of correlations against $(L^\infty, \mathcal{H}_\beta)$ observables. The ASIP holds for zero-mean observables in \mathcal{H}_β .*

1.6. Necessity of positive entropy for statistical quasistability. A smooth map $f \in \mathcal{F} \in \mathcal{F}_{\text{NSD}}$ is called *Collet-Eckmann* if for each critical value $v \in \mathcal{V}$, all forward iterates are defined and $\liminf_{n \rightarrow \infty} \frac{1}{n} \log |Df^n(v)| > 0$. A smooth interval map $f \in \mathcal{F} \in \mathcal{F}_{\text{NSD}}$ is called *Misiurewicz* if all periodic points are hyperbolic repelling and if, for all critical points c (in the standard sense that $Df(c) = 0$), the orbit of c avoids a neighbourhood of the critical set. By Mañé's theorem (see for example [MeS, Theorem III.5.1]) or [M1], any Misiurewicz map is also Collet-Eckmann. A smooth Collet-Eckmann interval map $f \in \mathcal{F} \in \mathcal{F}_{\text{NSD}}$ with non-flat critical points has an acip [CE].

Sometimes, given a piecewise-monotone map of an interval, a subinterval will be forward-invariant under some iterate of the map. One refers to this property as *renormalisability*, see [MeS, §II.5a] for definitions and details. One can study the dynamics restricted to such forward-invariant subintervals independently from the dynamics of points that never enter these subintervals. Repeated renormalisability entails a stratification of the phase space. As a potential varies, equilibrium states may jump from one level to another in a non-smooth fashion, as considered in [D1]. This does not happen for non-renormalisable maps under reasonable transitivity assumptions.

Our next theorem goes in the other direction to Corollary 1.7: even if we restrict to the well-behaved class of non-renormalisable smooth Misiurewicz maps, we do not have statistical quasistability of acips without (for example) the uniform entropy condition, once we leave the topological class. Contrast with [Ru5], which gives linear response when one remains within the class and suggested it may hold if one remains tangential to the class.

Theorem 1.30 (Varying entropy implies instability). *Let $\mathcal{F} \in \mathcal{F}_{\text{NSD}}$ denote the class of C^2 unimodal maps f with non-positive Schwarzian derivative, having $f(\partial I) \subset \partial I$ and $|Df|_{\partial I}| > 1$, endowed with the C^0 topology. Consider a continuous one-parameter family of maps $(f_t)_{t \in [0, \theta]} \subset \mathcal{F}$, for some $\theta > 0$. Suppose that f_0 is non-renormalisable and has all periodic points repelling, and that the topological entropy of f_t is not locally constant at $t = 0$. Let $C > 0$ and suppose that for all $t \in [0, \theta)$, $\int_I \log |Df_t(x)| dx > -C$. Let p be a non-boundary periodic point of f_0 .*

Then there is a sequence $t_k \rightarrow 0$ for which each f_{t_k} is a non-renormalisable Misiurewicz map with acip μ_k and for which the measures μ_k converge to the equidistribution on the orbit of p . Moreover, $h(\mu_k) \rightarrow 0$.

Since any f_0 -invariant probability measure (not supported on ∂I) can be approximated by equidistributions on periodic orbits (noting f_0 is Misiurewicz, this is easy to show), there are

t_k such that the μ_k converge to whatever f_0 -invariant probability measure we please. That the entropy tends to zero follows from Theorem 1.18.

As a remark, the condition on the integral of $\log |Df_t(x)|$ just says that the critical points are uniformly not too flat. It is clearly satisfied by non-trivial quadratic maps of the interval. See [BM, D2] for results concerning existence of acips under this condition and non-existence when it fails. The theorem immediately implies Theorem 1.8.

1.7. Positive entropy does not imply statistical stability. In this section we state that equilibrium states do not vary continuously (in general), even if their entropy is bounded away from zero.

Theorem 1.31. *There exists a sequence $(f_k)_{k \geq 1} \in \mathcal{F}_{\text{NSD}}$ having decreasing critical relations with $f_k \rightarrow f_0$ as $k \rightarrow \infty$ with the following properties. Each f_k , $k \geq 0$ has an acip μ_k with entropy uniformly bounded away from 0; the measures μ_k converge to a strictly convex combination of the equilibrium state μ_0 for f_0 and a Dirac mass on a repelling fixed point.*

The sequence of maps we construct (in §16.2) can be considered a toy model for a first return map to an interval containing the critical point of a unimodal map. Such a first return map will typically have one unimodal branch and infinitely many non-critical branches. Our toy model has one expanding branch and one unimodal branch. We expect that in the quadratic family similar discontinuity of equilibrium states will hold.

1.8. Background. For many classes of uniformly hyperbolic dynamical systems, statistical stability is well-established (in this paper we will think of a dynamical system as the dynamics $f : I \rightarrow I$, plus some associated measure μ on I). In particular, stability holds for systems where the dynamics $f : I \rightarrow I$ has an underlying (finite) Markov structure and the measures μ_f are equilibrium states for Hölder potentials $\varphi_f : I \rightarrow \mathbb{R}$. Indeed, for many such systems, it is known that the measures do not merely vary continuously with the map, but actually vary differentiably. For example, if $f_t : M \rightarrow M$ are C^3 Axiom A diffeomorphisms of a manifold M , each with a unique *physical* measure μ_t , and the family $t \mapsto f_t$ is C^3 , then the map $t \mapsto \int \psi d\mu_t$ is differentiable at $t = 0$ for any real-analytic observable $\psi : M \rightarrow \mathbb{R}$, see [KKPW, Co, Ru3]. This theory has been further developed, placing it inside a general theory of ‘linear response’ for dynamical systems, see example [Ru4] and references therein, as well as comments on the work of Baladi and Smania below.

1.8.1. Physical measures. An invariant measure is said to be *physical* if there is a positive (full-dimensional) Lebesgue measure set of points $x \in M$ for which the average of Dirac measures along the orbit of x converges to the invariant measure. An ergodic acip is a physical measure, as is the equidistribution on the orbit of a periodic attractor. For smooth unimodal maps, physical measures can be very strange. For example, a quadratic map with a physical measure which is the Dirac mass on a repelling fixed point was constructed in [BK]. Perhaps even more surprisingly, topologically transitive, unimodal maps of high critical order can have their physical measure supported on an absorbing Cantor set [BKNS].

1.8.2. Acips for smooth interval maps. For example, let \mathcal{F}_Q be the class of quadratic interval maps $f_a : x \mapsto ax(1-x)$, for $0 < a \leq 4$, and consider the potential $\varphi(x) := -\log |Df|$ for $f \in \mathcal{F}_Q$. By [BL], Lebesgue measure is ergodic. Any acip is a physical measure. Moreover it is an equilibrium measure with respect to the potential φ , this following from [L] and [P].

On the other hand, if the map is hyperbolic, so it has a (hyperbolic) attracting periodic orbit, then the (physical measure and) equilibrium state is the equidistribution on the attracting orbit. Hyperbolic parameters form an open, dense subset of $(0, 4]$ ([Ly, GS]), yet the set of parameters for which f_a has an acip, by [J, BC], has positive Lebesgue measure. There are further parameters for which equilibrium states do not exist. This gives an indication of the complexity of the possible behaviour within the quadratic family.

There are positive results regarding statistical stability in \mathcal{F}_{NSD} . In [Ts2, F, FT] it was shown that for large families $\mathcal{F} \in \mathcal{F}_{\text{NSD}}$ of smooth maps with acips, the acips depend continuously on the map. In the first two of these papers, the maps had to have some exponential growth along the critical orbit, while in the final one, subexponential growth was sufficient. However in all cases it was essential that the growth constants which defined \mathcal{F} had to be uniform in the family. For the purposes of the current paper, we should point out that these uniform constants lead to uniform tail estimates (which correspond in our setting to $T_{\Delta}^R = 0$, see §11, and a uniform lower bound of the entropy of the acips for all maps in \mathcal{F}).

1.8.3. Prior results for fixed maps. For a positive measure set of quadratic parameters near $a = 4$, Pesin and Senti showed existence of equilibrium states and analyticity of the pressure for t in a neighbourhood of $[0, 1]$ and the potential $-t \log |Df|$ [PSe]. In [BT, IT1], these results were extended to eventually give existence of equilibrium states for $t \in (t^-, t^+)$ and differentiability of the pressure function for transitive (multimodal) maps. Recently, in [PR1], existence of equilibrium states and analyticity of the pressure were proven by Przytycki and Rivera-Letelier. Ideas were further developed in [GPR]. The paper [PR1] transferred some results of [PRS, PR2] for rational maps to the interval setting. In the same paper, strong results concerning alternative definitions of pressure and concerning non-uniform hyperbolicity (Topological Collet-Eckmann condition, etc.) were proven. They also show statistical properties such as decay of correlations and the Central Limit Theorem. Compared with that paper, ours has a very different (and, we feel, more canonical) method of constructing induced maps. We focus more on convergence properties of measures and thermodynamic quantities, while they concentrate on the pressure. Our results on almost upper-semicontinuity of the free energy are surprising. We have slightly weaker (see §1.8.4) hypotheses too, not requiring critical points to be non-flat, and allowing discontinuities and indifferent points. Of course, our paper is geared towards proving results about families of maps; that we obtain both new and (slight generalisations of) recent results for fixed maps is a bonus.

1.8.4. Non-positive Schwarzian derivative and bounded distortion. For the proof of Theorem 1.18, we need distortion bounds and lower bounds on the derivative, see Lemma 10.3. Uniformity of the bounds for the sequence of maps is important and follows from non-positive Schwarzian derivative and the uniform extendibility (*aka* Koebe space) for the induced maps. If we assume the conclusions of Lemma 10.3 hold for our sequences, we can prove Theorem 1.18. For further results, we just need some iterate of our induced maps to be expanding—the uniformity of Lemma 10.7 is unnecessary (N, K may depend on f) for our purposes. Expansion follows from uniform bounded distortion for iterates of the induced maps. For fixed maps, therefore, we could just assume some distortion bounds as done in [PR1, Definition 1.10]; for sequences of maps we would need to assume uniformity in these bounds. For simplicity we assume non-positive Schwarzian derivative. We choose non-positive rather than negative because if one considers an infinitely-renormalisable quadratic map, the limit maps in renormalisation theory [Su, Theorem 1] automatically have non-positive Schwarzian derivative, and this will be useful in applications of this work.

1.8.5. *Return maps in the Hofbauer extension.* First return maps in the Hofbauer extension [H2] were introduced in [B] and used more recently to good effect in [BT, IT1], for example. However, those works just look at the first return to a single interval, or ‘extendible column’ in the extension, which engenders full-branched induced Markov maps. The tail estimates for such maps, however, are quite weak and led to difficulties in applications.

An important technical advance in this work is to take return maps to a possibly large, but finite, union of intervals. Combinatorial estimates then give strong control on the tails of the corresponding induced maps, see Lemma 4.9.

For convergent sequences of maps, Hofbauer extensions need not converge. This problem was defined away in [FT] by excluding some possible limit maps, including any map with a preperiodic critical point. We overcome this difficulty by embedding our return maps in an ambient space, where a subsequence of the return maps converges. This limit return map will be an induced map for the limit base map, but may not correspond to a return map in the limit map’s Hofbauer extension.

1.8.6. *Upper-semicontinuity of entropy.* In piecewise-monotone families, topological entropy is lower-semicontinuous (note the number of branches is fixed). Misiurewicz has produced examples which show one cannot do better than lower-semicontinuous [M2], see also [ALM, Section 4.5]. Measures of maximal entropy exist and have the same entropy as the map [H1]. Therefore metric entropy is not upper-semicontinuous in general (free energy is neither, then). In [MS], conditions are given for topological entropy to be continuous at a map (in terms of periodic orbits containing singularities). They consider all convergent sequences; we limit ourselves to those with decreasing critical relations but obtain more limit points, so there is some but not complete overlap. We prove upper semi-continuity of the metric entropy. This additionally implies that topological entropy is continuous at the limit map (for that sequence of maps).

1.8.7. *Keller’s example.* In [K1], Keller considers a sequence of continuous expanding, piecewise linear maps of the interval with four branches in the form of a W. The two external branches can be assumed to be full. The internal branches meet at a tip just above the diagonal, and have slope approximately $2 - \varepsilon$. The fixed point near the tip and its symmetric preimage define a restrictive interval on which the map is a tent map with slope $2 - \varepsilon$. The acip for the map is supported on the restrictive interval. It has entropy equal to log of the slope of the internal branches, thus bounded away from zero. In the limit (where $\varepsilon \rightarrow 0$), the tip meets the diagonal and the restrictive interval vanishes. The limit of the acips is an atom on the tip; meanwhile the limit map has its acip. This shows lack of quasistability. It can occur because the limit map has an extra critical relation: the turning point at the tip becomes a fixed point. This example has been further developed in [LGBPE, EM], where all maps are transitive; the limit map still has an extra critical relation. On the other hand, the examples of Theorems 1.30, 1.31 have decreasing critical relations.

1.8.8. *Failure of statistical stability for acips for smooth interval maps.* Tsujii showed that, in the quadratic family, the Chebyshev parameter 4 is accumulated by a positive measure set of once-renormalisable Collet-Eckmann parameters whose acips converge, as the parameters converge to 4, to the Dirac mass on 0 [Ts2, Remark 1.2]. These ideas were further developed by Thunberg [Th]. In particular, statistical stability does not hold for any full measure set of Collet-Eckmann parameters in the quadratic family. This lack of stability was due to renormalisation.

In Theorem 1.30, we show that statistical stability of acips does not hold even if one restricts to the very well-behaved, non-renormalisable Misiurewicz parameters (or in Corollary 1.9, if one restricts to full-measure sets of non-renormalisable Collet-Eckmann parameters). At the boundary of Tsujii's and Thunberg's renormalisation intervals are parabolic maps. Developing ideas of Homburg and Young [HomY], we find non-renormalisable and almost-parabolic Misiurewicz maps whose measures approximate the equidistribution along the parabolic orbit. Showing convergence of measures is somewhat technical, requiring us to keep track of constants in the proof of existence of the acips, unlike say in the related estimates of Benedicks and Misiurewicz in [BM]. The entropy of the acips along our sequences of maps tends to zero. The proof of the theorem occupies §15. In Theorem 1.31, we show that a lower bound on entropy does not guarantee statistical stability.

1.8.9. *Stronger versions of statistical stability.* In many works, see for example [RS, Al, AV, AS] the convergence in $L^1(m)$ of the densities $\frac{d\mu_f}{dm}$ of acips μ_f is called *strong statistical stability*. The notion of statistical stability of measures we use here is sometimes referred to as weak statistical stability. One can also ask for even stronger results. Baladi, in [Ba1, §1], see also [Ba2, §3.2], asks about the smoothness of measure-parameter dependence for Collet-Eckmann unimodal maps. Our Theorem 1.30 implies that some sort of uniformity condition on the Collet-Eckmann constants is required for continuity (let alone smoothness). Baladi and Smania, in [BaS1], showed that the physical measure μ_0 (an acip) for a tent map f_0 actually depends differentiably if the family of tent maps $(f_t)_{t \in (-\varepsilon, \varepsilon)}$, with corresponding physical measures $(\mu_t)_{t \in (-\varepsilon, \varepsilon)}$, is chosen to be tangent to the topological class of f_0 . On the other hand, for unimodal maps with critical points, the corresponding result in [BaS2] (see also [Ru5]) requires that the maps stay in a fixed topological class. That requirement implied (see Appendix A of [BaS2]) uniformity of the Collet-Eckmann constants, as well as constant topological entropy. While this paper was being prepared, a paper of Baladi, Benedicks and Schnellmann appeared proving sharp results on statistical stability for a subset of the Collet-Eckmann parameters [BBS]. These are stronger conclusions, but for a much less general class of maps than those considered here.

1.8.10. *Higher dimensions.* For results on statistical stability for non-uniformly hyperbolic maps in higher dimensions, see [V] and [ACF] where partially hyperbolic diffeomorphisms and Hénon maps are considered. See also [Al, Ar]. The results in the current paper are restricted to the one-dimensional case, but we expect that some analogue of our results carries over to the higher-dimensional case, probably requiring *sufficiently large* entropy as in [Bu] – the main difficulty is then to build families of Markov extensions with the right properties for this setting.

1.9. **Acknowledgements.** Some of this work came out of discussions between MT and J.M. Freitas and M. Holland at the University of Porto. He would like to thank them for their comments and encouragement. He would also like to thank M. Sambarino for his question about convergence of pressure. We thank V. Baladi, F. Przytycki and J. Rivera-Letelier for several insightful conversations.

2. METHOD AND STRUCTURE

First we give an overview of how almost-semicontinuity of the free energy may hold. The structure of the remainder of the paper is then presented. Finally we introduce our running assumption.

2.1. Method. Due to the presence of critical points, it is convenient to use the Hofbauer extension (\hat{I}, \hat{f}) of the original system (I, f) . It is locally homeomorphic. In Definition 4.8 we introduce the important (*canonical*) level- R induced map $(\hat{X}(R), \hat{F}, \tau)$, where $\hat{X}(R) \subset \hat{I}$ is a certain finite union of R -cylinders in the Hofbauer extension, \hat{F} is the (Markovian) first return map to $\hat{X}(R)$ with return time function τ . Note that $\hat{X}(R)$ is a finite union of intervals and that branches of the first return map will not be surjective.

Positive-entropy measures in \mathcal{M}_f lift to measures on the extension (\hat{I}, \hat{f}) . Counting arguments on the extension imply that if $R \geq 8d$ and $d \geq 2$,

$$\varepsilon(R) = \varepsilon(R, d) := 8 \frac{\log R}{R}, \quad (2.1)$$

and

$$\eta(R, d) := \frac{\varepsilon(R)^2}{2R(\log d)^2}, \quad (2.2)$$

then any measure $\mu \in \mathcal{M}_f$ with entropy at least $2\varepsilon(R)$ must lift to a measure $\hat{\mu}$ with

$$\hat{\mu}(\hat{X}(R)) \geq \eta(R, d).$$

The normalised restriction $\hat{\nu}$ of the measure $\hat{\mu}$ to $\hat{X}(R)$ is \hat{F} -invariant. Kac' Lemma implies that

$$\int \tau d\hat{\nu} \leq \frac{1}{\eta(R, d)}$$

(see [A, §1.5], for example, for a proof of both of these facts). The counting arguments of Lemma 4.6 are an important refinement of the ideas in [H2, Theorem 9], and better implemented than in [BT, Lemma 4].

For convergent sequences of d -branched maps, level- R induced maps need not converge to a level- R induced map for the limit system. Embedding the induced maps in some ambient space, some subsequence of the induced maps does converge to a well-defined limit map, with controlled properties.

The estimates hitherto depend only on R and d , so they hold for sequences of maps. We study convergence of induced measures $\hat{\nu}$ embedded into the ambient space. The limit induced measure will *spread* (Definition 8.5) to a light limit measure for the limit map (which may or may not be the one desired).

Tail estimates on the first return times then come into play. Our counting arguments imply that the number of branches with return time equal to n is exponentially small, bounded by $\exp(-n\varepsilon(R))$. This implies in some vague sense that the branches with high return time make little contribution to the entropy.

Drops in the Lyapunov exponent in the limit occur when diminishing mass for the induced measures has non-diminishing integral of $\log |DF|$, where F represents the induced maps. This bit of integral gets lost in the limit. This only happens if there is also a drop T_{Δ}^R in the integral of the return time.

If $T_{\Delta}^R = 0$ for some R , then the Lyapunov exponent is continuous (for the original sequence of measures). This also holds even if $\lim_{R \rightarrow \infty} T_{\Delta}^R = 0$. As entropy is upper-semicontinuous (Theorem 1.15), for these cases free energy will also be upper-semicontinuous. Otherwise, if the limit is positive, the spread of the limit induced measure will have entropy strictly greater than the limit of the entropies. Either this measure will have free energy greater than or equal to the limit or, to balance out, some bit of free energy gets lost in the limit. In this latter case, some

other similarly-constructed measure must have free energy strictly greater than the limit free energy, or one can approximate the lost free energy (a quantity greater than E_+) by a periodic orbit, giving a contradiction. Almost upper-semicontinuity of the free energy is obtained.

We induce F again to obtain a full-branched Markov map with (exponential) good tail estimates. This useful result is stated as Theorem 14.10. Application of thermodynamic formalism and ergodic theory for countable state Markov shifts gives analyticity of the pressure and uniqueness of equilibrium states, under transitivity assumptions. The almost sure invariance principle and decay of correlations follows, since our full-branched induced map can be viewed as a Young tower with exponential tails.

2.2. Structure. In §3, we detail some properties of $J(f)$. In §4, we introduce notions concerning the Hofbauer extension and its topological structure and we carry out some counting arguments in Lemma 4.6. Canonical level- R induced maps are introduced in Definition 4.8. In the subsequent section, we examine convergence of cylinder sets in Hofbauer extensions. In §6, we embed the Hofbauer extensions (and thus $\hat{X}(R)$) into an ambient space, obtaining convergence of the induced maps there to some limit map. Continuing on the topological side, we then show in §7 that given some orbit, there is another orbit which mimics the part of the orbit corresponding to large return times. The error term is uniform for a given sequence of maps.

Up until this point, everything was topological. Next, in §8, we describe the correspondence between measures on the interval and measures for induced maps. The lift of a positive entropy measure to the Hofbauer extension cannot have much mass near the boundary of $\hat{X}(R)$, we show. From this tightness-type result, we obtain limit induced measures in §9. Any drop in (metric) entropy is small compared to the (possible) drop T_Δ^R in the integral of the inducing time. Upper-semicontinuity of metric entropy is proven.

We present some standard distortion and expansion estimates for maps with non-positive Schwarzian derivative in §10. Theorem 1.18 is finally proven in §11, with a casewise analysis depending on the behaviour of T_Δ^R .

Next follows a section presenting some Katok theory (with self-contained, simple proofs). In §13 we prove several results depending on Theorem 1.18 and Katok theory.

In §14, facts concerning ergodic theory and thermodynamic formalism for countable state Markov shifts are reviewed. We obtain a full-branched induced map with exponential tails (with respect to the lift of the equilibrium measure). Thanks to the estimates for our first return maps defined earlier, thermodynamic formalism almost automatically gives analyticity of the pressure and uniqueness of equilibrium states. We show the almost sure invariance principle and decay of correlations.

In §15 we prove Theorem 1.30 which, we recall, shows that for any natural family of unimodal maps and corresponding measures, uniform positive entropy is necessary for quasistability.

Finally, in §16, we show, for a toy model, that uniform positive entropy *does not imply* stability of acips.

2.3. Running assumption. Given any d -branched piecewise-monotone map $f : \cup_{j=1}^d I_j \rightarrow I$, we know $J(f)$ is forward-invariant and $J(f) \cap \mathcal{E}(f) = \emptyset$. In Lemma 3.1 we show that any positive-entropy measure $\mu \in \mathcal{M}_f$ must have $\mu(J(f)) = 1$. Therefore to prove the results stated in this introduction, we can and do, without loss of generality, make the following assumption.

In the definition of piecewise-monotone maps, each I_j is open.

Also without loss of generality, **we assume** $|I| = 1$.

3. PIECEWISE PRELIMINARIES

Lemma 3.1. *Given a piecewise-monotone map f , without the running assumption that each I_j is open, the following holds. Any positive-entropy measure $\mu \in \mathcal{M}_f$ must have $\mu(J(f)) = 1$.*

Proof. To show this, note first that μ is ergodic with positive entropy. Hence it must be nonatomic, so any countable set has measure zero. Moreover, μ is invariant so almost every point is recurrent. It suffices, therefore, to show that almost every point is accumulated on both sides by points from $\cup_{n \geq 0} f^{-n}(\mathcal{E})$. We can, moreover, assume I is open.

Consider the set

$$E = I \setminus \overline{\bigcup_{n \geq 0} f^{-n}(\mathcal{E})}.$$

It is an open set, so its connected components consist of a countable number of intervals. Note that E is forward-invariant, $f(E) \subset E$. Fix some connected component E_j of E . If $f^n(E_j) \cap E_j \neq \emptyset$ for some $n > 0$, then $f^n(E_j) \subset E_j$. Either some iterate of f is not defined anywhere on E_j or all iterates are defined and homeomorphic on E_j . Either way, it follows from ergodicity (measures in \mathcal{M}_f are ergodic), invariance and positive entropy that $\mu(E_j) = 0$. The boundary points of the components E_j form a countable set, so they have measure zero. Let E_* be the union of the closures of the connected components of E . Thus $I \setminus E_*$ has full measure. But this is exactly the set of points accumulated on both sides by points from $\cup_{n \geq 0} f^{-n}(\mathcal{E})$. We conclude that almost every point is in $J(f)$. \square

Uniqueness of equilibrium measures will rely on the following rather general lemma. Its proof is very similar to that of Lemma 5.2 of [D2]. Here we assume transitivity; in the setting of [D2], ergodicity was assumed, giving transitivity on a set of full measure.

Lemma 3.2. *Let f be a piecewise-monotone map, transitive on $J(f)$. Suppose there is an open interval W containing a point from $J(f)$ and $l \geq 1$ such that $W \subset f^l(W)$. Then there exists $N \geq 1$ such that*

$$J(f) \subset \cup_{j=1}^N f^j(W).$$

Proof. We can assume, without loss of generality, that $l = 1$, since f^l would also be a piecewise-monotone map. Let

$$W_j = \bigcup_{i=0}^j f^i(W) = f^j(W).$$

Since $\mathcal{E} = \mathcal{E}(f)$ is finite, there is an N such that for all $j > N$, $W_j \setminus W_N$ does not form a one-sided neighbourhood of any point of \mathcal{E} .

Suppose $j \geq N$ and V is a connected component of $W_j \setminus \mathcal{E}$. Suppose there is some $k \geq 1$ such that $f^k(V) \cap W_N \neq \emptyset$ and let l be the least such k . Then $f^l|_V$ is continuous.

For each connected component V of $W_N \setminus \mathcal{E}$, let n_V be the minimal $k \geq 1$ such that $f^k(V) \cap W_N \neq \emptyset$ if such a k exists; otherwise set $n_V = 0$. Let M be the maximum of the n_V , noting that W_N has a finite number of connected components. Let X be a connected component of W_j for some $j \geq M$. Suppose X does not contain a connected component of W_M . Let $y \in X$ and let $l \geq 1$ be minimal such that $y \in f^l(V)$, where V is some component of $W_N \setminus \mathcal{E}$. Then $l > M$. In particular, $n_V = 0$, so $f^i(y) \notin W_N$ for all $i \geq 0$. This holds for each $y \in X$. Therefore $f^i(X) \cap W_N = \emptyset$ for all $i \geq 0$. By transitivity, $X \cap J(f) = \emptyset$.

Let V_1, \dots, V_r denote the connected components of W_M . For $j \geq M$ and $1 \leq k \leq r$, let V_k^j denote the connected component of W_j containing V_k . These are the only connected components of W_j which may contain points from $J(f)$.

Let $W_\infty = \bigcup_{j \geq 0} W_j$. We shall say that a point x has the *one-sided property* if W_∞ contains nested, one-sided neighbourhoods Z_k of the x such that: each Z_k contains points from $J(f)$, each $Z_k \not\subset W_j$ for any $j < \infty$, and $|Z_k| \rightarrow 0$ as $k \rightarrow \infty$.

Note that if the conclusion of the lemma does not hold, there exists a point x which has the one-sided property. Suppose this is so. We must arrive at a contradiction.

Points with the one-sided property belong to the finite set

$$\bigcup_{k=1}^r \partial \bigcup_{j \geq M} V_k^j.$$

If y is in the interior of W_∞ , then $f^k(y)$ does not have the one-sided property for any $k \geq 0$. Suppose x has the one-sided property and let Z_x be a corresponding one-sided neighbourhood of x . Then there is a (we can assume strictly monotone) sequence $(y_n)_n \subset J(f)$ for which $f(y_n) \in Z_x$ and $f^n(y) \rightarrow x$ as $n \rightarrow \infty$. Let y be its limit. Then $f((y_n, y))$ contains $(f(y_n), x) \subset Z_x$. Since x has the one-sided property, $(y_n, y) \not\subset W_j$ for any $j > 0$, so y has the one-sided property. Moreover, if $y_n \in U_i$ for all large n , then the continuous extension of f to \overline{U}_i maps y to x .

It follows that (the finite collection of) points having the one-sided property are periodic, in the sense that if x, Z_x are as before and Z_x is sufficiently small then some iterate $f^k(Z_x)$ is a small one-sided neighbourhood of x . The point x is a fixed point of the continuous extension g of $f|_{Z_x}$ to the closure $\overline{Z_x}$. If $g(Z_x) \subset Z_x$ for arbitrarily small Z_x then $J(f)$ must just be the orbit of x , by transitivity, and thus be contained in $f^k(W)$. Otherwise, x is a (one-sided, topologically-) repelling fixed point for g . For points of $J(f)$ to accumulate near x in Z_x , they have to come from somewhere: there must be a point $y \neq x$ and arbitrarily small, one-sided neighbourhoods Z_y of y mapped by f^k onto one-sided neighbourhoods of x in Z_x , with Z_y containing points from $J(f)$. But y, Z_y are not periodic in the above sense, so y cannot have the one-sided property. In particular, some sufficiently small J_y is contained in some W_j . Then $W_{j+k} \supset f^k(J_y)$ contains some J_x , contradicting the one-sided property. \square

4. COMBINATORICS IN THE HOFBAUER EXTENSION

We describe the dynamically defined cylinders. Let $\mathcal{P}_0 = \mathcal{P}_0^f := I$ and $\mathcal{P}_1 = \mathcal{P}_1^f := \{I_1, \dots, I_d\}$. For $n \geq 2$, let

$$\mathcal{P}_n = \mathcal{P}_n^f := \bigvee_{i=0}^{n-1} f^{-i}(\mathcal{P}_1).$$

Each $\mathbf{C} \in \mathcal{P}_n$ is an n -cylinder: a maximal open interval on which f^n is defined and $f^n : \mathbf{C} \rightarrow f^n(\mathbf{C})$ is a homeomorphism. Its boundary points are in the set $\cup_{j \geq 0} f^{-j}(\mathcal{E})$. We let $\mathbf{C}_n[x]$ denote the member of \mathcal{P}_n containing x .

Lemma 4.1. *For $x \in J(f)$, $\mathbf{C}_n[x]$ is (unambiguously) defined for all $n \geq 0$ and contains x in its interior and $\mathbf{C}_n[x]$ shrinks to the point x as $n \rightarrow \infty$.*

Proof. This follows immediately from the definition of $J(f)$. □

One can code the cylinder sets as follows. For $n \geq 1$, let

$$\Sigma_d^n := \{1, 2, \dots, d\}^n.$$

Any $\omega \in \Sigma_d^n$ is of the form $\omega = (\omega_1 \omega_2 \dots \omega_n)$. Given such an ω , there exists at most one n -cylinder, denoted \mathbf{C}^ω , such that $f^{l-1}(\mathbf{C}^\omega) \subset I_{\omega_l}$ for $l = 1, \dots, n$. Of course, each cylinder is of this form for some ω .

Set

$$\mathcal{D} = \{f^k(\mathbf{C}) : k \geq 0, \mathbf{C} \in \mathcal{P}_k\}.$$

As \mathcal{D} is a set, each element $D \in \mathcal{D}$ appears once. As in [H2], the *Hofbauer extension* is defined as

$$\hat{I} := \bigsqcup_{D \in \mathcal{D}} D.$$

We call each D a *domain* of \hat{I} . There is a natural projection map $\pi : \hat{I} \rightarrow I$. A point $\hat{x} \in \hat{I}$ is in some $D \in \mathcal{D}$; \hat{x} can be represented by (x, D) where $x = \pi(\hat{x})$. Given $\hat{x} \in \hat{I}$, we can denote the domain $D \in \mathcal{D}$ it belongs to by $D_{\hat{x}}$.

The map

$$\hat{f} : \pi^{-1} \left(\bigcup_{j=1}^d I_j \right) \rightarrow \hat{I}$$

is defined by

$$\hat{f}(\hat{x}) = \hat{f}(x, D) = (f(x), D')$$

if there are cylinder sets $\mathbf{C}' \subset \mathbf{C}$, with $\mathbf{C}' \in \mathcal{P}_{k+1}$ and $\mathbf{C} \in \mathcal{P}_k$, such that

$$x \in f^k(\mathbf{C}') \subset f^k(\mathbf{C}) = D$$

and $D' = f^{k+1}(\mathbf{C}')$. Equivalently, there are k, y such that $f^k(y) = x$, $f^k(\mathbf{C}_k[y]) = D$ and $f^{k+1}(\mathbf{C}_{k+1}[y]) = D'$. In this case, we write $D \rightarrow D'$, giving $(\mathcal{D}, \rightarrow)$ the structure of a directed graph. A path from D_1 to D_2 in $(\mathcal{D}, \rightarrow)$ is a sequence of arrows in the graph leading from D_1 to D_2 . The map π acts as a semiconjugacy between \hat{f} and f :

$$\pi \circ \hat{f} = f \circ \pi.$$

The Hofbauer extension has the useful property of being Markovian:

Lemma 4.2. *Let $\hat{x} \in D \in \mathcal{D}$, let $n \geq 1$ and suppose $D' \in \mathcal{D}$ contains $\hat{f}^n(\hat{x})$. Then some neighbourhood $W \subset D$ of \hat{x} is mapped homeomorphically by \hat{f}^n onto D' .*

Proof. This follows by induction from the definition of \hat{f} . \square

Lemma 4.3. *Let f be a piecewise-monotone map. If $\hat{x}_1, \hat{x}_2 \in \hat{I}$ have $\pi(\hat{x}_1) = \pi(\hat{x}_2) \in J(f)$ then there exists $n \in \mathbb{N}$ such that $\hat{f}^n(x_1) = \hat{f}^n(x_2)$.*

Proof. Suppose that $x = \pi(\hat{x}_1) = \pi(\hat{x}_2)$ and let D_1 and D_2 denote the domains in the Hofbauer extension containing \hat{x}_1 and \hat{x}_2 respectively. By Lemma 4.1, there exists an $n \geq 1$ such that $\mathbf{C}_n[x] \subset \pi(D_1 \cap D_2)$. The construction of the extension implies that $\hat{f}^n(\hat{x}_1), \hat{f}^n(\hat{x}_2) \in \mathbf{C}_n[x]$ and hence that $\hat{f}^n(\hat{x}_1) = \hat{f}^n(\hat{x}_2)$. \square

Lemma 4.4. *Suppose f is transitive on $J(f)$. Let $D_1, D_2 \in \mathcal{D}$ be domains for which the first return maps each have at least two branches. Then there is a path from D_1 to D_2 .*

Proof. The first return map to any domain in \mathcal{D} is full-branched, by Lemma 4.2, so D_1 and D_2 each contain infinitely many repelling periodic points. Meanwhile, there are only finitely many periodic points in the post-critical orbits, that is, contained in the set $\cup_{n \geq 0} f^n(\mathcal{V})$. Therefore there are infinitely many repelling periodic points in each of D_1 and D_2 which project down to points in $J(f)$. Let $\hat{p}_1 \in D_1$ and $\hat{p}_2 \in D_2$ be such points, with projections p_1 and p_2 respectively.

Let $W_1 \subset \pi D_1$ be an open interval containing p_1 , small enough that $W_1 \subset f^l(W_1)$. By Lemma 3.2, there is some N for which $\cup_{j=0}^N f^j(W_1)$ contains p_2 . Thus some $\hat{y} \in D_1$ is mapped by some \hat{f}^j to $\pi^{-1}(p_2)$. By Lemma 4.3, some further iterate of \hat{y} by \hat{f}^j is actually equal to \hat{p}_2 . \square

We denote the *base* of \hat{I} , the copy of I in \hat{I} , by D_0 . The natural inclusion map sending I to D_0 we denote by $\iota : I \hookrightarrow \hat{I}$. For $D \in \mathcal{D}$, we define $\text{lev}(D)$ to be the length of the shortest path $D_0 \rightarrow \dots \rightarrow D$ starting at the base D_0 .

Definition 4.5. *For each $R \in \mathbb{N}$, let \hat{I}^R be the bounded part of the Hofbauer extension defined by*

$$\hat{I}^R := \sqcup \{D \in \mathcal{D} : \text{lev}(D) \leq R\}.$$

We define $\hat{\mathcal{P}}_n$ to be the set of intervals $\{\pi|_D^{-1}\mathbf{C} : D \in \mathcal{D}, \mathbf{C} \in \mathcal{P}_n\}$. For a domain $D \in \mathcal{D}$, let D_ℓ^R be the left-most element of $\hat{\mathcal{P}}_R$ in D and let D_r^R be the rightmost. Set

$$\hat{I}_-^R := \hat{I}^R \cap \sqcup_{D \in \mathcal{D}} D \setminus (D_\ell^R \cup D_r^R).$$

As we will see later, excluding the boundary components will give uniform distortion bounds (thanks to the Koebe principle) for certain induced maps. Boundary components project to subsets of R -cylinders.

For the purposes of the following lemma, let us set $\mathcal{P}_n^0 := \iota^{-1}\mathcal{P}_n$, the same collection of n -cylinders except living on the base of the extension, and for a word ω , we set $\mathbf{C}_0^\omega := \iota^{-1}(\mathbf{C}^\omega)$.

Recall from (2.1) that $\varepsilon(R) = 8 \log R/R$.

Lemma 4.6. *Let $R \geq 8d$. Let*

$$\delta := \frac{\varepsilon(R)}{4 \log d} = \frac{2 \log R}{R \log d}.$$

There exists $n_0 = n_0(R)$ such that for any piecewise-monotone d -branched map f and any $n \geq n_0$, the following holds. Let \mathcal{Q} denote the collection of elements of \mathcal{P}_{nR}^0 which visit \hat{I}_-^R at most δn times in the first nR iterates. Then

$$\#\mathcal{Q} \leq R^{7n} < \exp(nR\varepsilon(R)).$$

Moreover, for $n \geq n_0$, the number of elements of \hat{P}_{R+n} contained in \hat{I}_-^R with first return time to \hat{I}_-^R equal to n is bounded by $\exp(n\varepsilon(R))$.

Proof. From the definition of $\varepsilon(R)$, $R^{7n} < \exp(nR\varepsilon(R))$.

By construction of the extension, if ω has length n , \hat{f}^j is homeomorphic on \mathbf{C}_0^ω for each $j \leq n$ (in particular, for each j , $\hat{f}^j(X)$ is contained in a single domain of the extension).

We shall count elements of \mathcal{P}_{nR}^0 for large n . The following estimates are needed.

- (i) if $X \in \mathcal{P}_l^0$ for some $l \geq 1$ then, for $1 \leq k \leq R$, there are at most two elements X' of \mathcal{P}_{l+k}^0 for which both $X' \subset X$ and $\hat{f}^{l+k}(X') \subset \hat{I} \setminus \hat{I}^R$;
- (ii) there are at most $(2dR)^2$ domains D with $\text{lev}(D) \leq R$; in particular, there are at most $8d^2R^2$ connected components of $\hat{I}^R \setminus \hat{I}_-^R$;
- (iii) $\binom{n}{\lfloor \delta n \rfloor} \leq 2^n$.

Estimate (i) is so because if X' does not have a common boundary point with X , then $\pi \hat{f}^l(X') \in \mathcal{P}_k$, so $\hat{f}^{l+k}(X') \subset \hat{I}^k \subset \hat{I}^R$. For estimate (ii), the projection of each boundary point of a domain D of level at most R is contained in the set $\cup_{0 \leq j \leq R-1} f^j(\mathcal{V})$ of cardinality bounded by $2dR$, since the set \mathcal{V} of critical values has cardinality at most $2d$. A domain has two boundary points, giving $(2dR)^2$ possibilities. The trivial estimate (iii) on the binomial coefficient could be improved (for example using Stirling's formula), but is sufficient for our needs.

To each element $\mathbf{C} = \mathbf{C}_0^\omega$ corresponds a word ω of length nR . For $j \geq 0$, denote by \mathbf{C}^j the image $\hat{f}^j(\mathbf{C})$, so $\pi \mathbf{C}^j$ is contained in the 1-cylinder of f corresponding to the symbol ω_{j+1} . We can divide the word ω into n blocks of length R . If $\mathbf{C}^{kR+j} \subset \hat{I} \setminus \hat{I}_-^R$ for $j = 0, \dots, R-1$ then (counting from zero) we say the k^{th} block is *external*. If it is external, there is a maximal $l \leq R$ for which $\mathbf{C}^{kR+j} \not\subset \hat{I}^R$ for $j = 0, \dots, l-1$ (so if $\mathbf{C}^{kR} \not\subset \hat{I}^R$ then $l = 0$, and if $l < R$ then $\mathbf{C}^{kR+l} \subset \hat{I}^R \setminus \hat{I}_-^R$). Given $\omega_1 \dots \omega_{kR}$, by (i) there are at most two possibilities for the sequence $\omega_{kR+1} \dots \omega_{kR+l}$. Given $\omega_1 \dots \omega_{kR+l}$, there are by (ii) at most $8d^2R^2$ possibilities for $\omega_{kR+l+1} \dots \omega_{(k+1)R}$ (one possibility for each boundary component). There are R possible values for l .

Multiplying these estimates, given $\omega_1 \dots \omega_{kR}$, there are at most $16d^2R^3$ possible words $\omega_{kR+1} \dots \omega_{(k+1)R}$ for which the k^{th} block is external. For general blocks of length R we use a crude estimate: there are at most d^R possible words. There are fewer than 2^n possible configurations of $\lfloor \delta n \rfloor$ general and $n - \lfloor \delta n \rfloor$ external blocks, by (iii).

Elements of \mathcal{Q} visit \hat{I}_-^R at most δn times in the first nR iterates. A word corresponding to an element of \mathcal{Q} has at least $n - \lfloor \delta n \rfloor$ external blocks of length R . Combining the estimates from

the preceding paragraph, we get an upper bound of

$$\#\mathcal{Q} \leq 2^n (16d^2 R^3)^n d^{R\delta n}. \quad (4.1)$$

Replacing δ by $\frac{2 \log R}{R \log d}$, we obtain

$$\#\mathcal{Q} \leq (32d^2 R^3)^n R^{2n}.$$

Since $R \geq 8d$,

$$\#\mathcal{Q} \leq R^{7n}.$$

The first statement is now proven.

To finish, we claim that the number of elements of $\hat{\mathcal{P}}_{R+n}$ contained in \hat{I}_-^R with first return time to \hat{I}_-^R equal to n is bounded by

$$(2dR)^2 (16d^2 R^3)^{\lfloor n/R \rfloor} d^{1+2R}.$$

We assembled this bound from the following components. An element of $\hat{\mathcal{P}}_{R+n}$ is defined by a word of length $n + R$ plus the starting domain. There are, as before, at most $(2dR)^2$ domains of level at most R . There are d possible choices for the first symbol of the word. Then there follows $\lfloor n/R \rfloor$ full external blocks, each having $16d^2 R^3$ possibilities, followed by a final $\leq 2R$ free symbols with at most d^{2R} possibilities.

Taking n large, we can ignore the terms not being raised to the n , and recalling $R \geq 8d$ we obtain a bound of

$$(R^5)^{n/R} = \exp(5n(\log R)/R) < \exp(n\varepsilon(R)),$$

completing the proof. \square

Lemma 4.7. *If $\hat{Y}, \hat{Y}' \in \hat{\mathcal{P}}_n$ and $\hat{f}(\hat{Y}) \cap \hat{Y}' \neq \emptyset$, then $\hat{f}^{-1}(\hat{Y}') \cap \hat{Y}$ is mapped homeomorphically by \hat{f} onto \hat{Y}' .*

Proof. This follows from the cylinder structure and Lemma 4.2. \square

Definition 4.8. *The (canonical) level- R induced map is the system $(\hat{X}(R), \hat{F}, \tau)$ and consists of: $\hat{X} = \hat{X}(R)$, the union of elements of $\hat{\mathcal{P}}_R$ contained in \hat{I}_-^R ; the first return map \hat{F} to \hat{X} ; the first return time function τ on $\hat{X}(R)$.*

The large range of \hat{F} , with multiple connected components, allows one to obtain powerful counting estimates.

Lemma 4.9. *The level- R induced map $(\hat{X}(R), \hat{F}, \tau)$ has the following properties.*

- (i) \hat{F} is defined on a countable union of pairwise-disjoint open subintervals \hat{X}_i of \hat{X} ;
- (ii) \hat{F} maps each \hat{X}_i homeomorphically onto a connected component of \hat{X} ;
- (iii) the return time τ is constant on each \hat{X}_i , thus we can denote it τ_i ;
- (iv) for each i , there is some $\hat{X}'_i \supset \hat{X}_i$ mapped homeomorphically by \hat{f}^{τ_i} onto the domain of \hat{I} containing $\hat{F}(\hat{X}_i)$.

If $n_0 = n_0(R)$ is given by Lemma 4.6, then

$$\#\{i : \tau_i = n\} \leq \exp(n\varepsilon(R))$$

for all $n \geq n_0$.

Proof. Properties (i)-(iv) follow from Lemma 4.7 and Lemma 4.2. The counting statement follows immediately from Lemma 4.6, since each \hat{X}_i is an $(R + \tau_i)$ -cylinder. \square

Let (\hat{X}, \hat{F}, τ) be a level- R induced map.

Lemma 4.10. *Let $N \geq 1$ and let \hat{Y}_j , $j \geq 1$, be the connected components of the domain of \hat{F}^N . Let $n_0 = n_0(R)$ be given by Lemma 4.9. Let ρ_j denote the value on \hat{Y}_j of the N^{th} return time to \hat{X} for \hat{f} , so $\hat{f}^{\rho_j} = \hat{F}^N$ on \hat{Y}_j . Then there is an n_1 (dependent on $N, d, R, \varepsilon(R), n_0$ but not on f) for which, for all $n \geq n_1$,*

$$\#\{j : \rho_j = n\} \leq e^{3n\varepsilon(R)}.$$

Proof. Let us write ε for $\varepsilon(R)$. With the notation of Lemma 4.9, set $K_k := \#\{i : \tau_i = k\}$ for each $k \geq n_0$, so $K_k \leq e^{k\varepsilon}$. There is a uniform bound K_* (depending just on d, R, n_0) on K_k for $k \leq n_0$. Then

$$\#\{j : \rho_j = n\} \leq \sum \prod_{i=1}^N K_{k_i}$$

where the sum is taken over all N -tuples k_1, \dots, k_N which sum to n . The product is then bounded by $\exp(2n\varepsilon)$, for $n \geq \log K_*/\varepsilon$. There are fewer than n^N summands, so if n is big enough that $n\varepsilon \geq N \log n$, then

$$\#\{j : \rho_j = n\} \leq n^N \exp(2n\varepsilon) \leq \exp(3n\varepsilon).$$

\square

Lemma 4.11. *Let ρ be the inducing time (the R^{th} return time) corresponding to \hat{F}^R . Let $K \geq 1$ and suppose $\rho(\hat{x}) \leq K$ for some $\hat{x} \in \hat{X}$. Let \hat{y} be in the boundary of the connected component of \hat{X} containing \hat{x} . Then (\hat{x}, \hat{y}) contains a $(K + R)$ -cylinder for \hat{f} .*

Proof. Let $k \geq R$ be minimal such that $\hat{f}^k(\hat{x}) \in \hat{X}$. Then $\rho(\hat{x}) \geq k$. If $\hat{f}^k \in D \in \mathcal{D}$, we can pull back the boundary cylinders D_l^R, D_r^R to obtain $(k + R)$ -cylinders separating x from $\partial\hat{X}$. \square

The map \hat{F} is typically not onto. Since \hat{X} has only a finite number of connected components and \hat{F} is Markovian, we can find subsets with good mixing properties. Let us say that connected components \hat{Y}, \hat{Y}' are *linked* if some iterate of \hat{F} maps \hat{Y} onto \hat{Y}' and some iterate also maps \hat{Y}' onto \hat{Y} . Being linked is clearly a symmetric and transitive relation.

We denote by $\hat{X}_{\mathcal{T}}$ the collection of all connected components \hat{Y} of \hat{X} for which the first return map from \hat{Y} to \hat{Y} has at least two branches. Given $\hat{Y} \in \hat{X}_{\mathcal{T}}$, we can write $\mathcal{Y}_{\hat{Y}}$ for the union of all components linked to \hat{Y} .

Definition 4.12. *We call $\mathcal{Y}_{\hat{Y}}$ the primitive component of \hat{X} containing \hat{Y} .*

Remark 4.13. *The number of connected components of \hat{X} is bounded by $(2dR)^2 d^R$ and this is trivially an upper bound on the number of primitive components of \hat{X} .*

Lemma 4.14. *If f is transitive on $J(f)$ then \hat{X} has only one primitive component.*

Proof. Let $\hat{Y}_1, \hat{Y}_2 \in \hat{X}_{\mathcal{T}}$, so their first return maps each have at least two branches. The return maps are full-branched, by Lemma 4.9, so \hat{Y}_1 and \hat{Y}_2 each contain infinitely many repelling periodic points. Meanwhile, there are only finitely many periodic points in the post-critical

orbits, that is, contained in the set $\cup_{n \geq 0} f^n(\mathcal{V})$. Therefore there are infinitely many repelling periodic points in each of \hat{Y}_1 and \hat{Y}_2 which project down to points in $J(f)$. Let $\hat{p}_1 \in \hat{Y}_1$ and $\hat{p}_2 \in \hat{Y}_2$ be such points, with projections p_1 and p_2 respectively.

Let $W_1 \subset \pi\hat{Y}_1$ be an open interval containing p_1 , small enough that $W_1 \subset f^l(W_1)$. By Lemma 3.2, there is some N for which $\cup_{j=0}^N f^j(W_1)$ contains p_2 . Thus some $\hat{y} \in \hat{Y}_1$ is mapped by some \hat{f}^j to $\pi^{-1}(p_2)$. By Lemma 4.3, some further iterate of \hat{y} by \hat{f}^j is actually equal to \hat{p}_2 . Repeating the argument switching twos and ones, we find a point in \hat{Y}_2 mapped to \hat{Y}_1 . It follows that \hat{Y}_1 and \hat{Y}_2 are linked, so they are subsets of the same primitive component, as required. \square

Lemma 4.15. *Let \mathcal{Y} be a primitive component of \hat{X} . Then the first return (under \hat{f}) map $\hat{F}_{\mathcal{Y}}$ to \mathcal{Y} has the following properties:*

- (i) $\hat{F}_{\mathcal{Y}}$ is defined on a countable union of pairwise-disjoint open subintervals \hat{X}_i of \mathcal{Y} ;
- (ii) $\hat{F}_{\mathcal{Y}}$ maps each \hat{X}_i homeomorphically onto a connected component of \mathcal{Y} ;
- (iii) there exists $N_{\mathcal{Y}}$ such that, for each \hat{X}_i ,

$$\bigcup_{j=1}^{N_{\mathcal{Y}}} \hat{F}_{\mathcal{Y}}^j(\hat{X}_i) \supset \mathcal{Y};$$

- (iv) the return time is constant on each \hat{X}_i and coincides with the return time under \hat{F} to \hat{X} : thus we can denote it τ_i ;
- (v) for each i , there is some $\hat{X}'_i \supset \hat{X}_i$ mapped homeomorphically by \hat{f}^{τ_i} onto the domain of \hat{I} containing $\hat{F}_{\mathcal{Y}}(\hat{X}_i)$.

If $n_0 = n_0(R)$ is given by Lemma 4.6, then

$$\#\{i : \tau_i = n\} \leq \exp(n\varepsilon)$$

for all $n \geq n_0$.

Proof. We must prove (iii) and (iv). The other statements follow from Lemma 4.9.

To show (iii), note first that $\hat{F}_{\mathcal{Y}}(\hat{X}_i)$ is a connected component of \mathcal{Y} . Each pair $\hat{Y}_1, \hat{Y}_2 \subset \mathcal{Y}$ of connected components of \mathcal{Y} is linked, so for some minimal $n = n(\hat{Y}_1, \hat{Y}_2)$,

$$\hat{F}_{\mathcal{Y}}^n(\hat{Y}_1) \supset \hat{Y}_2.$$

There are only a finite number of such ordered pairs, so

$$N_{\mathcal{Y}} := 1 + \max_{\hat{Y}_1, \hat{Y}_2 \in \mathcal{Y}} n(\hat{Y}_1, \hat{Y}_2)$$

is finite and has the desired property.

Now we show (iv). Suppose that \hat{x} is in the domain of $\hat{F}_{\mathcal{Y}}$, so $\hat{x} \in \hat{Y}_1$ and $\hat{F}_{\mathcal{Y}}(\hat{x}) \in \hat{Y}_2$ for some connected components \hat{Y}_1, \hat{Y}_2 of \mathcal{Y} . Let \hat{Y}' denote the connected component of \hat{X} containing $\hat{F}(\hat{x})$. In particular, $\hat{F}(\hat{Y}_1) \supset \hat{Y}'$. Some \hat{F}^j , $j \geq 0$, maps \hat{Y}' onto \hat{Y}_2 . Since \hat{Y}_2, \hat{Y}_1 are linked, some iterate of \hat{F} maps \hat{Y}_2 onto \hat{Y}_1 . Consequently \hat{Y}_1 and \hat{Y}' are linked. Hence $\hat{Y}' = \hat{Y}_2 \in \mathcal{Y}$ and

$$\hat{F}(\hat{x}) = \hat{F}_{\mathcal{Y}}(\hat{x}) = \hat{f}^{\tau_i}(\hat{x}),$$

if $\hat{F} = \hat{f}^{\tau_i}$ on the interval $\hat{X}_i \ni \hat{x}$. \square

Definition 4.16. A level- R primitive induced map is such a triplet $(\mathcal{Y}, \hat{F}_{\mathcal{Y}}, \tau)$, where \mathcal{Y} is a primitive component of a level- R induced map, $\hat{F}_{\mathcal{Y}}$ is the first return map to \mathcal{Y} and τ is the return time (w.r.t. \hat{f}).

5. CODING CONVERGENCE

Recall that the coding of cylinder sets \mathbf{C}^ω , for $\omega \in \Sigma_d^n$, was defined near the start of §4. Introducing f -dependence, we can write $\mathbf{C}^\omega(f)$ for \mathbf{C}^ω . Let $\Sigma^n(f)$ denote the set of $\omega \in \Sigma_d^n$ for which $\mathbf{C}^\omega(f)$ exists. Sequences of maps $(f_k)_k$ in what follows will be sequences of d -branched piecewise-monotone maps.

Lemma 5.1. Let $(f_k)_k$ be a sequence converging to f_0 as $k \rightarrow \infty$ and having decreasing critical relations. Given $n \geq 1$, for all sufficiently large k , $\Sigma^n(f_0) = \Sigma^n(f_k)$. For $\omega \in \Sigma^n(f_0)$,

$$\mathbf{C}^\omega(f_k) \rightarrow \mathbf{C}^\omega(f_0)$$

in the Hausdorff metric as $k \rightarrow \infty$.

Proof. Since the f_k converge to f_0 and cylinder sets are open intervals, $\Sigma^n(f_0) \subset \Sigma^n(f_k)$ for large k and the corresponding cylinder sets converge. To complete the proof, we prove by contradiction that $\Sigma^n(f_k) \subset \Sigma^n(f_0)$ for all large k .

The case $n = 1$ is trivial, by definition of convergence of the f_k . So suppose $n \geq 2$ is minimal such that for some subsequence $(k_i)_i$, tending to infinity with i , there is some ω with

$$\omega \in \Sigma^n(f_{k_i}) \setminus \Sigma^n(f_0)$$

for all i . To simplify notation, we can suppose that this subsequence is the original sequence $(f_k)_k$. Denote by ω' the word ω with its first symbol removed. Then ω' is a word of length $(n - 1)$ and $f_k(\mathbf{C}^\omega(f_k)) \subset \mathbf{C}^{\omega'}(f_k)$. Since n is minimal, $\omega' \in \Sigma^{n-1}(f_0)$. It follows that $f_0(I_{\omega_1}(f_0))$ shares a common boundary point, x say, with $\mathbf{C}^{\omega'}(f_0)$, and that

$$f_0(I_{\omega_1}(f_0)) \cap \mathbf{C}^{\omega'}(f_0) = \emptyset.$$

Let c_0 be the boundary point of $I_{\omega_1}(f_0)$ whose small one-sided neighbourhoods in $I_{\omega_1}(f_0)$ get mapped to a one-sided neighbourhood of x . Let $p \geq 0$ be minimal such that $f_0^p(x) \in \mathcal{E}(f_0)$ and note that f_0^p is a homeomorphism on some neighbourhood of x . Therefore on any small neighbourhood W of x , f_k^p is a homeomorphism for all k large enough and the corresponding boundary points $x_k \in \partial \mathbf{C}^{\omega'}(f_k)$ to x lie in W . There is a critical relation of order $(p + 1)$ between c_0 and $f_0^p(x)$ for f_0 . By the decreasing critical relations hypothesis, this critical relation also exists for f_k . In particular, $\{x_k\} = W \cap \partial f_k(I_{\omega_1}(f_k))$, for W small and k sufficiently large. But this implies that

$$f_k(I_{\omega_1}(f_k)) \cap \mathbf{C}^{\omega'}(f_k) = \emptyset$$

so $\mathbf{C}^\omega(f_k) = \emptyset$, a contradiction. \square

Lemma 5.2. Let $(f_k)_k$ be a sequence converging to f_0 as $k \rightarrow \infty$ and having decreasing critical relations. For each j and n with $0 \leq j \leq n$, for each $\omega \in \Sigma^n(f_0)$ and for all k large enough, the sets

$$f_k^j(\mathbf{C}^\omega(f_k))$$

exist and they converge in the Hausdorff metric to (the non-trivial set) $f_0^j(\mathbf{C}^\omega(f_0))$ as $k \rightarrow \infty$.

Proof. This follows from convergence of the maps and Lemma 5.1. \square

Lemma 5.3. *Let $(f_k)_k$ be a sequence converging to f_0 as $k \rightarrow \infty$ and having decreasing critical relations. For each j and n with $1 \leq j, n$, there is a $\kappa > 0$ such that the following holds. For each $\omega \in \Sigma^j(f_0)$ and $\theta \in \Sigma^n(f_0)$, for all k large enough, either $\mathbf{C}^\omega(f_k) \cap f_k^n(\mathbf{C}^\theta(f_k)) = \emptyset$ or*

$$|\mathbf{C}^\omega(f_k) \cap f_k^n(\mathbf{C}^\theta(f_k))| \geq \kappa.$$

Proof. For the purposes of the proof, we may assume that $\mathbf{C}^\omega(f_k) \cap f_k^n(\mathbf{C}^\theta(f_k))$ is non-empty. Then the set

$$W_k = \mathbf{C}^\theta(f_k) \cap f_k^{-n}(\mathbf{C}^\omega(f_k))$$

is non-empty and is an $(n+j)$ -cylinder. By Lemma 5.1, W_k converges to an $(n+j)$ cylinder W_0 for f_0 . From Lemma 5.2, $f_k^n(W_k)$ converges to $f_0^n(W_0)$. Taking $\kappa = |f_0^n(W_0)|/2$ completes the proof. \square

Notation. Recall that for a map f , $\hat{X} = \hat{X}(R)$ is the union of elements of \hat{P}_R contained in \hat{I}_R^- . The first return map to \hat{X} we denote \hat{F} ; its return time we denote τ . Given maps f_k and a number $R \geq 1$, we introduce dependence on k to the notation \mathcal{P} , \hat{P}_n , \hat{I}^R , $\hat{X}(R)$, \hat{X} , \hat{F} , τ , giving the notation \mathcal{P}^k , \hat{P}_n^k , $\hat{I}^R(k)$, $\hat{X}^k(R)$, \hat{X}^k , \hat{F}^k , τ_k .

Lemma 5.4. *Let $(f_k)_k$ be a sequence converging to f_0 as $k \rightarrow \infty$ and having decreasing critical relations. For each $R, K \geq 1$, there is a $\kappa > 0$ such that the following holds.*

$$\inf_k \{ |V| : V \in \hat{\mathcal{P}}_K^k, V \subset \hat{I}^R(k) \} \geq \kappa > 0.$$

Proof. This follows straightforwardly from Lemma 5.3 with $j = K$ and $n = R$. \square

6. EXTENSION EMBEDDING

Given d, R , we can order the set $\cup_{j \leq R} \Sigma_d^j$. Then for a d -branched map f , we can embed \hat{I}^R naturally into

$$I_* := I \times \bigcup_{j \leq R} \Sigma_d^j$$

as follows. To each $\hat{x} \in \hat{I}^R$, there is a minimal word

$$\theta = \theta(\hat{x}) \in \bigcup_{j \leq R} \Sigma_d^j$$

for which projection of the domain of \hat{I}^R containing \hat{x} is $f^{L(\theta)}(\mathbf{C}^\theta(f))$. Embed \hat{x} as $(x, \theta(\hat{x}))$, and let us call the embedding map ι_R . Recall we denote by $(\hat{X}(R), \hat{F}, \tau)$ the level- R induced map.

Definition 6.1. *We call $(X(R), F, \tau)$ the embedded level- R induced map, where $X = X(R)$ is the embedding of $\hat{X} = \hat{X}(R)$, and F is the corresponding map, so*

$$F = \iota_R \circ \hat{F} \circ \iota_R^{-1}.$$

The domain of definition of τ is extended by $\tau|_X := \tau \circ \iota_R^{-1}$.

We can write π_1 for projection onto the first coordinate. Then $\pi_1 \circ F = f^\tau \circ \pi_1$ on the domain of definition of F . Of course, F and τ are only defined on a subset of I_* . One should think of F as being Markovian, as it inherits the properties of \hat{F} given in Lemma 4.9.

To a branch (Z, θ) of F , there is a word ω corresponding to Z , such that $Z = \mathbf{C}^\omega(f)$, whose length $L(\omega)$ equals $\tau + R$ on (Z, θ) . There is another unique word θ_* corresponding to the domain containing $F((Z, \theta))$. The triplet $(\omega, \theta, \theta_*)$ entirely specifies the branch (Z, θ) , in the sense that it tells us the interval Z and the starting and image domains in I_* .

Let Ω denote the set of such triplets corresponding to branches of F . Let Ω^N denote the set of $(\omega, \theta, \theta_*) \in \Omega$ for which $L(\omega) \leq N$. Let Ω' be the (finite) set of pairs (ω', θ) , with $L(\omega') = R$, for which $(\mathbf{C}^{\omega'}, \theta)$ is a connected component of X . As suh,

$$X := \bigcup_{\omega' \in \Omega'} (\mathbf{C}^{\omega'}(f), \theta).$$

For $k \geq 1$ and f_k (in place of f), we introduce the notation $F_k, \tau_k, X^k, \Omega_k, \Omega_k^N$ and Ω'_k .

For each $N \in \mathbb{N}$, we shall say that the sequences $(\Omega_k^N)_k$ and (Ω'_k) are *eventually constant* if, for all k, k' large enough, $\Omega_k^N = \Omega_{k'}^N$ and $\Omega'_k = \Omega'_{k'}$. Suppose that for all $N \geq 1$, the sequences $(\Omega_k^N)_k$ and (Ω'_k) are eventually constant. We denote by Ω_0^N and Ω'_0 their respective limits. Set $\Omega_0 := \bigcup_N \Omega_0^N$. We define

$$Y^0 := \bigcup_{\omega' \in \Omega'_0} (\mathbf{C}^{\omega'}(f_0), \theta).$$

For $(\omega, \theta, \theta^*) \in \Omega_0$ and $(x, \theta) \in (\mathbf{C}^\omega(f_0), \theta)$, we define F_0 of (x, θ) by

$$F_0(x, \theta) := (f_0^{L(\omega)-R}(x), \theta_*)$$

and τ_0 by

$$\tau_0(x, \theta) = L(\omega) - R.$$

The domain and range of F_0 are subsets of Y^0 . Convergence of f_k to f_0 gives $F_0(x, \theta) = \lim_{k \rightarrow \infty} F_k(x, \theta)$ (noting that the branches are open sets). It makes sense to say that the sequence $((X^k, F_k, \tau_k))_k$ is convergent and we can write (still omitting the R -dependence)

$$(X^k, F_k, \tau_k) \rightarrow (Y^0, F_0, \tau_0). \quad (6.1)$$

We call (Y^0, F_0, τ_0) a (embedded) *level- R limit induced map*.

In general, for each N (and d, R), there are only a finite number of possibilities for Ω^N . Thus there is always a (strictly increasing) sequence $(n_k)_k$ for which for each $N \geq 1$, the subsequence $(\Omega_{n_k}^N)_{n_k}$ is eventually constant. Similarly, there is a subsequence for which $(\Omega'_{n_k})_{n_k}$ is eventually constant.

Remark 6.2. *If the map f_0 has a finite post-critical set $\bigcup_{n \geq 0} f^n(\mathcal{V})$, then the corresponding Hofbauer extension has only a finite number of domains. This need not be true for maps f_k converging to f_0 , so there is no particular reason the level- R limit induced map should coincide with the level- R induced map for f_0 . This is why we need to consider convergence of the embeddings.*

7. DYNAMICS OF RARE RETURNS

In this section we construct orbits with large inducing time, which corresponds to only returning infrequently to the bounded part \hat{X} in the Hofbauer extension.

Suppose we have a convergent sequence of embedded level- R induced maps

$$(X^k, F_k, \tau_k) \rightarrow (Y^0, F_0, \tau_0)$$

as per (6.1). For U, V connected components of X^k (or of Y^0 if $k = 0$), we can define $\tau_k(U, V)$ as follows. Let A be a subinterval of U for which there is an n_A with $F_k^{n_A}(A) = V$. Let $p_A = \tau + \tau \circ F_k + \dots + \tau \circ F_k^{(n_A-1)}$. Define $\tau_k(U, V)$ as the infimum over all such A of p_A . If $\tau_k(U, V) < +\infty$, we can choose an A which minimises p_A . Then we can define a *quick word*

$$b^k(UV) = Z_0 Z_1 \dots Z_{n_A-1},$$

where Z_j is the connected component of X^k containing $F_k^j(A)$.

Since Y^0 has a finite number of connected components,

$$M := \max\{\tau_0(U, V) : \tau_0(U, V) < +\infty\}$$

is finite.

Lemma 7.1. *For each K , for all large k the following holds. If $j \geq 1$, U, V are connected components of X^k , if $\tau_k(F_k^i(A)) \leq K$ for each $i = 0, 1, \dots, j-1$ and if F_k^j maps $A \subset U$ homeomorphically onto V , then $\tau_k(U, V) \leq M$.*

Proof. By convergence, for each pair of connected components U, V ,

$$\liminf_{k \rightarrow \infty} \{\tau_k(U, V) : \tau_k(U, V) > M\} = +\infty. \quad (7.1)$$

Consequently, for all k large enough and all connected components U, V, W of X^k , provided $\tau_k(U, V), \tau_k(V, W) \leq M$, one has $\tau_k(U, W) \leq 2M$ and so

$$\tau_k(U, W) \leq M. \quad (7.2)$$

Now let U, V, A be as per the hypotheses. For $i = 0, 1, \dots, j$, let V_i be the connected component of X^k containing $F_k^i(A)$. Let $l \leq j$ be the maximal integer for which for each $i = 0, \dots, l$, $\tau_k(U, V_i) \leq M$. If $l < j$ then $\tau_k(V_l, V_{l+1}) > M$ by (7.2) and so $\tau_k(V_l, V_{l+1}) > K$ by (7.1). This contradicts the assumption $\tau_k(F_k^l(A)) \leq K$, completing the proof of the lemma. \square

Let ξ_k denote the set of connected components of the domain of F_k ; let $\xi_k(K)$ denote those components on which $\tau_k \geq K$. Let

$$V_K^k := \bigcup_{Z \in \xi_k(K)} Z \subset X^k.$$

Let (without indicating the dependence on k) e_n be the n^{th} return time (with respect to F_k) to V_K^k and set

$$g(y) := \sum_{i=0}^{e_1(y)-1} \tau_k(F_k^i(y)).$$

Lemma 7.2. *Suppose $x \in V_K^k$ is recurrent.*

For arbitrarily large n , there is an n -periodic point $w \in V_K^k$ such that $F_k^{e_j(w)}(w)$ and $F_k^{e_j(x)}(x)$ are in the same element of $\xi_k(K)$ for $j = 0, \dots, n-1$ and such that if $\tau_k(F_k^{e_j(w)+1}(w)) < K$ then $g(F_k^{e_j(w)+1}(w)) \leq M$.

There is $y \in V_K^k$ such that $F_k^{e_n(y)}(y)$, and $F_k^{e_n(x)}(x)$ are in the same element of $\xi_k(K)$ for each $n \geq 0$ and such that if $\tau_k(F_k^{e_n(y)+1}(y)) < K$ then $g(F_k^{e_n(y)+1}(y)) \leq M$.

Proof. If $F_k^j(x) \in Z_j \in \xi_k$, we obtain a sequence $\underline{x} = Z_0 Z_1 Z_2 \dots$. Write V_j for the connected component of X^k containing Z_j . If $n, n+1, \dots, n+l$ satisfy

$$Z_j \in \xi_k \setminus \xi_k(K)$$

for $j = n, \dots, n+l$, and $Z_{n-1}, Z_{n+l+1} \notin \xi_k \setminus \xi_k(K)$, then we call $Z_n Z_{n+1} \dots Z_{n+l}$ a *low string*. Replacing each low string by the corresponding quick word $b^k(V_n V_{n+l})$, we obtain a new sequence $\underline{w} = W_0 W_1 W_2 \dots$. By construction, $F_k(W_j) \supset W_{j+1}$, while $W_0 \in \xi_k(K)$. Since x is recurrent, W_0 and W_n lie in the same connected component of X^k for infinitely many $n \rightarrow \infty$. Therefore, for a sequence of arbitrarily large n , there exists an n -periodic point $w = w_n$ such that $F_k^j(w) \in W_j$ for $j = 0, \dots, n$.

Taking a convergent sequence of w_n gives a limit point y which shadows x forever.

The estimates for g hold by construction. □

8. LIFTING AND SPREADING MEASURES AND ABRAMOV'S FORMULA

8.1. Lifts of measures. Given $\mu \in \mathcal{M}_f$, we say that μ *lifts to* \hat{I} if there exists an ergodic \hat{f} -invariant probability measure $\hat{\mu}$ on \hat{I} such that $\hat{\mu} \circ \pi^{-1} = \mu$. The process of constructing the lifted measure given in [K2] is as follows. First set $\hat{\mu}_0 := \mu \circ \iota^{-1}$. Now define $\hat{\mu}_n$ on \hat{I} by

$$\hat{\mu}_n := \frac{1}{n} \sum_{k=0}^{n-1} \hat{\mu}_0 \circ \hat{f}^{-k}. \quad (8.1)$$

Any vague limit of $(\hat{\mu}_n)_n$ is an \hat{f} -invariant measure. However, this limit might be the null measure. Keller showed that positive entropy is enough to guarantee existence of the lifted measure:

Lemma 8.1 ([K2, Theorem 3]). *Every positive entropy measure $\mu \in \mathcal{M}_f$ lifts to a measure $\hat{\mu} \in \mathcal{M}_{\hat{f}}$ with $\hat{\mu}(\hat{I}) = 1$ and the same entropy, $h(\hat{\mu}) = h(\mu)$. The sequence of measures $\hat{\mu}_n, n \geq 1$, defined in (8.1) converges weakly to $\hat{\mu}$.*

Recall that $\hat{X} = \hat{X}(R)$ denotes the union of elements of $\hat{\mathcal{P}}_R$ contained in \hat{I}_-^R . Note that $\hat{I}_-^R \setminus \hat{X}$ is a finite set which therefore has zero mass for any positive-entropy ergodic invariant probability measure. From (2.1), (2.2), $\varepsilon(R) = 8 \log R/R$, while

$$\eta(R, d) = \frac{\varepsilon(R)^2}{2R(\log d)^2}.$$

Lemma 8.2. *For any piecewise-monotone d -branched map f and any $\mu \in \mathcal{M}_f$ with entropy $h(\mu) \geq 2\varepsilon(R)$, the lift $\hat{\mu}$ of μ satisfies*

$$\hat{\mu}(\hat{X}) \geq \eta(R, d).$$

Proof. The following is a more detailed proof of the somewhat sketchy [BT, Lemma 4] which, moreover, yields more information: in particular uniform constants for families of maps together with a tail estimate.

As in Lemma 4.6, let \mathcal{Q} denote the collection of elements of \mathcal{P}_{nR}^0 which visit \hat{I}_-^R at most δn times in the first nR iterates, where $\delta = \varepsilon(R)/4 \log d$. By Lemma 4.6, for all large n , $\#\mathcal{Q} \leq \exp(nR\varepsilon(R))$. Of course, there is a trivial bound $\#\mathcal{P}_{nR} \leq d^{nR}$.

Now we can use these two estimates on the numbers of nR -cylinders to estimate entropy. For any finite sets of non-negative numbers $\{a_k\}_k, \{b_k\}_k$ such that $\sum_k a_k = a$, $\sum_k b_k = b$ and $a + b = 1$, Jensen's inequality implies that

$$-\sum_k a_k \log a_k - \sum_k b_k \log b_k \leq a \log \#\{a_k\}_k + b \log \#\{b_k\}_k + \log 2,$$

which we apply in the following entropy estimate. Since the partition \mathcal{P}_1 is generating for any positive entropy measure μ and since $\iota_{nR}^0 = \mathcal{P}_{nR}$, we deduce the bound

$$\begin{aligned} 2\varepsilon(R) \leq h(\mu) &\leq -\frac{1}{nR} \sum_{\mathbf{C} \in \mathcal{P}_{nR}} \mu(\mathbf{C}) \log \mu(\mathbf{C}) \\ &\leq -\frac{1}{nR} \left(\sum_{\mathbf{C} \in \mathcal{P}_{nR} \setminus \iota \mathcal{Q}} \mu(\mathbf{C}) \log \mu(\mathbf{C}) + \sum_{\mathbf{C} \in \iota \mathcal{Q}} \mu(\mathbf{C}) \log \mu(\mathbf{C}) \right) \\ &\leq \mu(\cup_{\mathbf{C} \in \mathcal{P}_{nR} \setminus \iota \mathcal{Q}} \mathbf{C}) \log d + \varepsilon(R) + \frac{\log 2}{nR}, \end{aligned}$$

for all large n . Therefore

$$\hat{\mu}^0(\cup_{\mathbf{C} \in \mathcal{P}_{nR}^0 \setminus \iota \mathcal{Q}} \mathbf{C}) = \mu(\cup_{\mathbf{C} \in \mathcal{P}_{nR} \setminus \iota \mathcal{Q}} \mathbf{C}) \log d \geq \varepsilon(R) - (\log 2)/nR.$$

In particular, for each large n , there is a set of $\hat{\mu}_0$ -measure at least $\varepsilon(R)/\log d - 1/nR = 8\delta - 1/nR$ which visits \hat{I}_-^R at least δn times in the first nR iterates. From the definition of $\hat{\mu}$, we get $\hat{\mu}(\hat{I}_-^R) = \hat{\mu}(\hat{X}) \geq 8\delta^2/R$. Substituting the definition of δ gives the required result. \square

Recall primitive components of \hat{X} were introduced in Definition 4.12.

Corollary 8.3. *For any $\mu \in \mathcal{M}_f$ with entropy $h(\mu) \geq 2\varepsilon(R)$, there is a primitive component \mathcal{Y}*

$$\hat{\mu}(\mathcal{Y}) \geq \eta(R, d).$$

Proof. From ergodicity and Definition 4.12, $\hat{\mu}$ can only give mass to one primitive component. By the lemma, this component has mass at least $\eta(R, d)$. \square

Given R and a measure $\mu \in \mathcal{M}_f$ with $h(\mu) \geq 2\varepsilon(R)$, then $\hat{\mu}(\hat{X}) \geq \eta(R, d)$ as in Lemma 8.2. We obtain a measure $\hat{\nu} = \hat{\nu}^R \in \mathcal{M}_{\hat{F}}$ defined by

$$\hat{\nu} := \frac{1}{\hat{\mu}(\hat{X})} \hat{\mu}|_{\hat{X}}, \tag{8.2}$$

since \hat{F} is a first return map. Recall the embedding ι_R from §6, giving (X, F, τ) , the embedded level- R induced map. The measure $\hat{\nu} \in \mathcal{M}_{\hat{F}}$ gives an *embedded induced measure* $\nu \in \mathcal{M}_F$ via $\nu = \hat{\nu} \circ \iota_R^{-1}$. In this setting, we have the following.

Lemma 8.4. *Given $\gamma > 0$, there exists $K \geq 1$ (K depending on d, R, γ) such that, if $\kappa := \min\{|V| : V \in \hat{\mathcal{P}}_K, V \subset \hat{I}^R\}$ then*

$$\nu(B(\partial X, \kappa)) < \gamma.$$

Proof. It suffices to show that

$$\hat{\nu}(B(\partial \hat{X}, \kappa)) < \gamma.$$

By Kac' Lemma, if ρ is the inducing time corresponding to \hat{F}^R , then

$$\int_{\hat{X}} \rho \, d\hat{\nu} = \frac{R}{\hat{\mu}(\hat{X})} \leq \frac{R}{\eta(R, d)}. \quad (8.3)$$

Let $K = R + \lceil R/(\eta\gamma) \rceil$, and let κ be given by the statement. By Lemma 4.11, $\rho \geq R/\eta\gamma$ on $B(\partial\hat{X}, \kappa)$. Thus by (8.3),

$$\hat{\nu}(B(\partial\hat{X}, \kappa)) < \gamma.$$

□

8.2. Spreading measures and Abramov's formula. Suppose we have some induced map (X, F, τ) , where (X, F, τ) is a level- R induced map, an embedded level- R induced map or a level- R limit induced map for an initial piecewise-smooth d -branched map f . Denote by π the projection from X to I . Let ν be an F -invariant probability measure, possibly non-ergodic. The pushforward $\pi_*\nu$ is a measure on I . Write ν^n for the restriction of ν to the set $\tau^{-1}(n)$ and put

$$T := \int \tau \, d\nu.$$

Definition 8.5. *The spread of ν is the measure*

$$\frac{1}{T} \sum_n \sum_{j=0}^{n-1} f_*^j(\pi_*\nu^n),$$

which is well-defined if $T < \infty$.

Writing μ for the spread of ν , Abramov's formula ([Ab1] or [Z]) states that

$$h(\mu) = \frac{1}{T} h(\nu). \quad (8.4)$$

If ν is ergodic, then straightforward application of Birkhoff's ergodic theorem gives

$$\int \log |Df| \, d\mu = \frac{1}{T} \int \log |DF| \, d\nu. \quad (8.5)$$

If ν is not ergodic, use ergodic decomposition: one can write $\mu = \int_{\mathcal{M}_f} \mu' \, dm_1(\mu')$ and

$$\nu = \int_{\mathcal{M}_f} \frac{T}{T_{\nu'}} \nu' \, dm_1(\mu')$$

for some probability measure m_1 and with μ' the spread of ν' . Then

$$\begin{aligned} \int \log |Df| \, d\mu &= \int_{\mathcal{M}_f} \int \log |Df| \, d\mu' \, dm_1(\mu') \\ &= \int_{\mathcal{M}_f} \frac{1}{T_{\nu'}} \int \log |DF| \, d\nu' \, dm_1(\mu') \\ &= \frac{1}{T} \int \log |DF| \, d\nu, \end{aligned}$$

so (8.5) still holds.

Naturally, given $(\hat{X}(R), F, \tau)$, if $\mu \in \mathcal{M}_f$ has entropy at least $2\varepsilon(R)$ with lift $\hat{\mu} \in \mathcal{M}_{\hat{F}}$, and $\hat{\nu} \in \mathcal{M}_{\hat{F}}$ is the normalised restriction to $\hat{X}(R)$ (which by Lemma 8.2 is well-defined), then $\hat{\nu}$ spreads to μ .

9. LIGHT LIMIT MEASURES AND ENTROPY

Let $(f_k)_k$ be a sequence of d -branched piecewise-monotone maps converging to f_0 as $k \rightarrow \infty$ and having decreasing critical relations. Let $R \geq 8d$. For each $k \geq 1$, let $\mu_k \in \mathcal{M}_{f_k}$ have $h(\mu_k) \geq 2\varepsilon(R)$ and suppose the sequence $(\mu_k)_k$ converges to some limit measure. Let

$$((X^k, F_k, \tau_k))_k \rightarrow (Y^0, F_0, \tau_0)$$

be a convergent sequence of level- R induced maps, as in §6. In particular, the sequences $(\Omega_k^N)_k$ are eventually constant for each N . We obtain the measures $\hat{\nu}_k \in \mathcal{M}_{\hat{F}_k}$ from (8.2), together with their embeddings $\nu_k \in \mathcal{M}_{F_k}$. Kac' Lemma says that

$$\int_{\hat{X}^k} \tau_k d\hat{\mu}_k = 1$$

so, writing $\eta = \eta(R, d)$, we have the following bound on the inducing time

$$\int \tau_k d\nu_k = \frac{1}{\hat{\mu}_k(\hat{X}^k)} \leq \frac{1}{\eta}. \quad (9.1)$$

Since the space $\overline{Y^0}$ is compact, there is a convergent subsequence of the measures ν_k whose limit is a probability measure supported on $\overline{Y^0}$. Let us suppose the sequence itself actually converges; the limit measure we shall denote $\nu_0 = \nu_0^R$. Its spread will be a light limit measure, as shown later in the section.

Lemma 9.1. *The measure ν_0 is F_0 -invariant.*

Proof. Let V be a connected component of $Y^0 \subset I_*$, so $\pi_1 V$ is an R -cylinder. Let U be an interval compactly contained in V such that

$$\nu_0(\partial U) = \nu_0(\partial(F_0^{-1}(U))) = 0,$$

that is, the boundaries of U and $F_0^{-1}(U)$ have zero measure. Since $\nu_0(\partial U) = 0$,

$$\lim_{k \rightarrow \infty} \nu_k(U) = \nu_0(U). \quad (9.2)$$

Now each connected component of $F_0^{-1}(U)$ is compactly contained in a branch of F_0 and so $\partial(F_0^{-1}(U) \cap \{\tau_0 \leq K\}) \subset \partial(F_0^{-1}(U))$ has zero measure for each K . For any K , since $\Omega_k^K = \Omega_0^K$ for all large k , the branches with inducing time bounded by K converge. Therefore,

$$\lim_{k \rightarrow \infty} \nu_k(F_k^{-1}(U) \cap \{\tau_k \leq K\}) = \nu_0(F_0^{-1}(U) \cap \{\tau_0 \leq K\}).$$

Meanwhile, (9.1) implies $\nu_k(\{\tau_k > K\}) < 1/K\eta$. Consequently (and again using the boundary has zero measure),

$$\lim_{k \rightarrow \infty} \nu_k(F_k^{-1}(U)) = \nu_0(F_0^{-1}(U)). \quad (9.3)$$

Since $\nu_k(U) = \nu_k(F^{-1}(U))$, (9.2)-(9.3) imply that $\nu_0(U) = \nu_0(F^{-1}(U))$. Approximating general sets compactly contained in V by open covers with massless boundary, one can then show that $\nu_0 = \nu_0 \circ F_0^{-1}$ on compact subsets of V and thus on V . It remains to treat boundaries of sets like V , that is, of R -cylinders.

So let Z^0 be an R -cylinder for f_0 and let $(Z^0, \theta) \subset I_*$. Let Z^k denote the corresponding cylinder for f_k , with a boundary point p_k such that p_k converges to a boundary point p_0 of Z^0 . Set $p'_k := (p_k, \theta)$. Let $\gamma > 0$ and let $K > R$ be given by Lemma 8.4. Let $\kappa > 0$ be given by Lemma 5.4. [Note that if $p'_k \notin \partial_{\mathcal{L}R} \hat{X}_k$ then $\nu_k(B(p'_k, \kappa)) = 0$.] Then Lemma 8.4 says that

$$\hat{\nu}_k(B(\partial \hat{X}^k, \kappa)) < \gamma$$

for all large k . In particular, $\nu_k(B(p'_k, \kappa)) < \gamma$. Thus $\nu_0(B(p'_0, \kappa/2)) \leq \gamma$. Letting $\gamma \rightarrow 0$, we conclude that the points projecting to boundaries of cylinders have zero ν_0 -measure, so $\nu_0 = \nu_0 \circ F^{-1}$ everywhere. \square

Let $T_k := \int \tau_k d\nu_k$; we will write T_k^R when we need to indicate dependence on R . Suppose that T_k converges to a finite quantity $T_L = T_L^R$, as $k \rightarrow \infty$. Let $T_0 = T_0^R := \int \tau_0 d\nu_0$ and set $T_\Delta^R := T_L^R - T_0^R$. Then

$$T_0 = \lim_{K \rightarrow \infty} \lim_{k \rightarrow \infty} \int_{\tau_k \leq K} \tau_k d\nu_k \leq T_L \leq \frac{1}{\eta}. \quad (9.4)$$

In particular, $T_\Delta^R \geq 0$. Since $\hat{\mu}_k(\hat{X}^k(R))$ is increasing with R , (9.1) implies that T_k^R is decreasing as R increases and

$$0 \leq T_\Delta^{R'} \leq T_L^{R'} \leq T_L^R \quad \text{for all } R' \geq R. \quad (9.5)$$

Lemma 9.2. *For some measure $\mu' \in \mathcal{M}_{f_0}$,*

$$\lim_{k \rightarrow \infty} \mu_k = \frac{T_0^R}{T_L^R} \mu_0^R + \left(1 - \frac{T_0^R}{T_L^R}\right) \mu',$$

where μ_0^R denotes the spread of ν_0 ($\nu_0 = \nu_0^R$).

Proof. Since the limit measure ν_0 is F_0 -invariant, it gives no mass to boundaries of cylinder sets. Denote by ν_k^n the restriction of ν_k to $\tau^{-1}(n)$. It follows that

$$\mu_{k,N} := \frac{1}{T_k} \sum_{n=1}^N \sum_{j=0}^{n-1} (f_k^j)_* (\pi_* \nu_k^n) \rightarrow \frac{1}{T_L} \sum_{n=1}^N \sum_{j=0}^{n-1} (f_0^j)_* (\pi_* \nu_0^n)$$

as $k \rightarrow \infty$. The difference between $\mu_{k,N}$ and μ_k tends to zero uniformly as $N \rightarrow \infty$. The right-hand side converges to $\frac{T_0^R}{T_L^R} \mu_0^R$ as $N \rightarrow \infty$, and the result follows. \square

Let us suppose in addition that $h(\nu_k^R)$ converges to a limit $h_L = h_L^R$ as $k \rightarrow \infty$. Then set

$$h_\Delta^R := h_L^R - h(\nu_0^R).$$

Lemma 9.3. *For all R with $\varepsilon(R) \leq \inf_k h(\mu_k)$,*

$$h_\Delta^R \leq 5\varepsilon(R)T_\Delta^R.$$

In particular, $\lim_{R \rightarrow \infty} (h_L^R - h(\nu_0^R)) \leq 0$.

Proof. We shall fix R and drop dependence on R from the notation. Let ξ_k denote the collection of branches of F_k . It is a generating partition for ν_k with entropy $h(\nu_k) < \infty$. Let

$$\xi_k^N := \bigvee_{j=0}^{N-1} F_k^{-j} \xi_k.$$

Suppose $h_\Delta > 0$, for otherwise there is nothing to prove, since $T_\Delta \geq 0$. Let $\gamma := h_\Delta/15$. Then there is an N for which

$$h(\nu_0) + \gamma > \frac{1}{N} H(\nu_0, \xi_0^N) := -\frac{1}{N} \sum_{Z \in \xi_0^N} \nu_0(Z) \log \nu_0(Z).$$

Let $\rho = \rho_k := \tau + \tau \circ F_k + \dots + \tau \circ F_k^{N-1}$, the N^{th} return time for F_k . Let $\xi_k^N(K)$ denote the elements of ξ_k^N on which $\rho \geq K$. Since $H(\nu_0, \xi_0^N)$ is finite, for each K large enough

$$-\frac{1}{N} \sum_{Z \in \xi_0^N(K)} \nu_0(Z) \log \nu_0(Z) < \gamma.$$

Then for large k ,

$$\begin{aligned} h(\nu_0) + \gamma &> -\frac{1}{N} \sum_{Z \in \xi_0^N \setminus \xi_0^N(K)} \nu_0(Z) \log \nu_0(Z) \\ &> -\frac{1}{N} \sum_{Z \in \xi_k^N \setminus \xi_k^N(K)} \nu_k(Z) \log \nu_k(Z) - \gamma \\ &\geq -\frac{1}{N} H(\nu_k, \xi_k^N) - \frac{1}{N} \sum_{Z \in \xi_k^N(K)} \nu_k(Z) \log \nu_k(Z) - \gamma \\ &\geq h(\nu_k) - \frac{1}{N} \sum_{Z \in \xi_k^N(K)} \nu_k(Z) \log \nu_k(Z) - \gamma. \end{aligned} \tag{9.6}$$

Passing to the second line required convergence of the measures on $\{\rho \leq K\}$. Rearranging and replacing $h(\nu_k)$ by (a lower bound) $h_L - \gamma$ produces

$$\frac{-1}{N} \sum_{Z \in \xi_k^N(K)} \nu_k(Z) \log \nu_k(Z) > h_\Delta - 3\gamma \tag{9.7}$$

for all sufficiently large k .

Thus we have a first upper bound for h_Δ . The exponential tails estimate of Lemma 4.10 will permit us to rewrite the upper bound in terms of ε and T_Δ .

Set

$$\beta_{n,k} := \nu_k(\{\rho = n\}).$$

By Kac' Lemma,

$$\lim_{k \rightarrow \infty} \sum_{n \geq 1} n \beta_{n,k} = \lim_{k \rightarrow \infty} NT_k = NT_L.$$

Hence $0 \leq \sum_{n \geq K} \beta_{n,k} \leq NT_k/K \leq (NT_L + 1)/K$ (for large k) which goes to 0 as $K \rightarrow \infty$. As a consequence,

$$\lim_{K \rightarrow \infty} \lim_{k \rightarrow \infty} \sum_{n \geq K} n \beta_{n,k} = NT_\Delta. \tag{9.8}$$

For $\varepsilon = \varepsilon(R)$, Lemma 4.10 says that for some n_1 independent of k , for all $n \geq n_1$,

$$\#\{Z \in \xi_k^N : \rho(Z) = n\} \leq \exp(3n\varepsilon).$$

Thus Jensen's inequality gives

$$-\frac{1}{N} \sum_{Z \in \xi_k^N : \rho(Z)=n} \nu_k(Z) \log \nu_k(Z) \leq \frac{\beta_{n,k} 3n\varepsilon}{N} - \frac{1}{N} \beta_{n,k} \log \beta_{n,k}. \tag{9.9}$$

Splitting the following sum over $n \geq K \geq n_1$ in two, depending on whether $\beta_{n,k}$ is greater or smaller than $e^{-n\varepsilon}$, one deduces that

$$\begin{aligned} \sum_{n \geq K} -\frac{1}{N} \beta_{n,k} \log \beta_{n,k} &\leq \frac{1}{N} \sum_{n \geq K} n \varepsilon \beta_{n,k} + \frac{1}{N} \sum_{n \geq K} e^{-n\varepsilon} n \varepsilon \\ &\leq \frac{1}{N} \sum_{n \geq K} n \varepsilon \beta_{n,k} + \frac{1}{K}, \end{aligned} \tag{9.10}$$

say, for all large K .

From (9.7), (9.9) and (9.10), for all K large enough,

$$\begin{aligned} h_\Delta - 3\gamma &< \sum_{n \geq K} \frac{\beta_{n,k} 3n\varepsilon}{N} - \frac{1}{N} \beta_{n,k} \log \beta_{n,k} \\ &\leq \frac{1}{K} + \frac{4\varepsilon}{N} \sum_{n \geq K} n \beta_{n,k}. \end{aligned}$$

Applying (9.8) and letting $K \rightarrow \infty$, $h_\Delta - 3\gamma \leq 4\varepsilon T_\Delta$. Since $\gamma = \varepsilon h_\Delta / 15$, we obtain

$$h_\Delta \leq 5\varepsilon T_\Delta, \tag{9.11}$$

as required \square

Proof of Theorem 1.15. Now let us prove upper semi-continuity of entropy for convergent sequences of measures $\mu_k \in \mathcal{M}_{f_k}$ with limit μ_∞ , where (f_k) is a convergent sequence of d -branched maps with decreasing critical relations. We may assume $h(\mu_k)$ converges to a strictly positive limit, for otherwise the result is trivial. Take R_0 large enough that $\varepsilon(R_0)$ is less than this limit. Passing to a subsequence if necessary, construct a convergent sequence of embedded level- R induced maps for each $R \geq R_0$. Obtain ν_k^R, T_k^R as before. Assume without loss of generality that ν_k^R converges to ν_0^R , $h(\nu_k^R)$ to h_L^R and T_k^R to T_Δ^R , for all $R \geq R_0$. Define h_Δ^R, T_Δ^R as before. Let μ_0^R denote the spread of ν_0 . By Lemma 9.2, (8.4) and Lemma 9.3 consecutively,

$$\begin{aligned} h(\mu_\infty) &\geq \frac{T_0^R}{T_L} h(\mu_0^R) \\ &= \frac{1}{T_L} h(\nu_0^R) \\ &\geq \frac{1}{T_L} h_L^R - 5\varepsilon(R) T_\Delta^R \\ &= \lim_{k \rightarrow \infty} h(\mu_k) - 5\varepsilon(R) T_\Delta^R. \end{aligned}$$

This holds for all large R . As $R \rightarrow \infty$, $\varepsilon(R) \rightarrow 0$ while, by (9.5), T_Δ^R remains bounded, so

$$h(\mu_\infty) \geq \lim_{k \rightarrow \infty} h(\mu_k)$$

as required. \square

10. NON-POSITIVE SCHWARZIAN DERIVATIVE

In this section we derive and use the distortion results required for our main theorems.

Definition 10.1. *If $g : Z \rightarrow g(Z)$ is a diffeomorphism on an interval Z , we say that the Minimum Principle holds if, for any $a, b \in Z$ and $x \in (a, b)$,*

$$|Dg(x)| \geq \min\{|Dg(a)|, |Dg(b)|\}.$$

For diffeomorphisms g with non-positive Schwarzian derivative, $|Dg|^{-1/2}$ is convex, so the Minimum Principle holds. To see a nice treatment of the following in our C^2 setting, see [C, Chapter 3].

Lemma 10.2 ([M1, p18, p24]). *Let $Z \subset Y$ be intervals and let g be a C^2 diffeomorphism from Y to $g(Y)$ with non-positive Schwarzian derivative. Extendibility implies bounded distortion: if $\kappa > 0$ and $\text{dist}(g(Z), \partial g(Y)) \geq \kappa|g(Z)|$, then for all $x, x' \in Z$,*

$$\frac{|Dg(x)|}{|Dg(x')|} \leq \left(\frac{1 + \kappa}{\kappa} \right)^2.$$

Lemma 10.3. *Let $(f_k)_k$ be a sequence converging to f_0 as $k \rightarrow \infty$, having decreasing critical relations and non-positive Schwarzian derivative. Let (X^k, \hat{F}_k, τ_k) be the corresponding level- R induced maps. For each R , there is a $C > 1$ such that*

$$\left| \frac{D\hat{F}_k^n(\hat{x})}{D\hat{F}_k^n(\hat{x}')} \right| \leq C$$

for each n, k and each \hat{x}, \hat{x}' in the same branch of \hat{F}_k^n . In particular, there is a uniform lower bound on $|D\hat{F}_k|$ independent of k .

Proof. Between any point of $\hat{X}^k(R)$ and the boundary of the element of \mathcal{D} containing the point, there lies an element of \hat{P}_R^k of length at least $\kappa > 0$, by Lemma 5.4. Lemma 4.2 and the standard Koebe distortion bound Lemma 10.2 will then imply the first statement. The second follows from the first since the components of \hat{X}^k have length at least κ . \square

Lemma 10.4. *Let $(\mathcal{Y}, \hat{F}_{\mathcal{Y}}, \tau)$ be a primitive level- R induced map. There is a $C > 0$ such that*

$$\lambda_{\min}\tau - C \leq \log |D\hat{F}_{\mathcal{Y}}| \leq \lambda_{\max}\tau + C.$$

Proof. Let Y_1, Y_2 be connected components of \mathcal{Y} . Since \mathcal{Y} is primitive, Y_1, Y_2 are linked. There is a minimal $C_{1,2} \geq 1$ for which there exists a subinterval V of Y_1 mapped by some iterate $\hat{f}_{\mathcal{Y}}^n$ onto Y_2 with derivative bounded inside $[C_{1,2}^{-1}, C_{1,2}]$ and with $n \leq C_{1,2}$. Let C' be the maximum of $C_{1,2}$ over all pairs Y_1, Y_2 of connected components of \mathcal{Y} .

Now let \hat{X}_i be a branch of $\hat{F}_{\mathcal{Y}}$, with $\tau = \tau_i$ on \hat{X}_i . For some n with $0 \leq n \leq C'$, $\hat{f}^{n+\tau_i}$ maps a subinterval $W \subset \hat{X}_i$ onto \hat{X}_i with

$$\log |D\hat{f}^{n+\tau_i}|_W - \log C' - C'' \leq \log |D\hat{F}_{\mathcal{Y}}|_{|\hat{X}_i} \leq \log |D\hat{f}^{n+\tau_i}|_W + \log C' + C'',$$

where C'' comes from bounded distortion, and $\tau \leq n \leq C + \tau$. There is a repelling periodic point in W , necessarily with exponent in $[\lambda_{\min}, \lambda_{\max}]$. Taking $C = C'\lambda_{\max} + \log C' + C''$, the result follows. \square

Lemma 10.5. *Let $\kappa > 0$. Let $Z \subset Y \subset g(Z)$ be intervals, where g is a C^2 diffeomorphism from Y to $g(Y)$ with non-positive Schwarzian derivative. Suppose $\text{dist}(g(Z), \partial g(Y)) \geq \kappa|g(Z)|$. Then at the fixed point p , $|Dg(p)| \geq \sqrt{1 + \kappa}$.*

Proof. Suppose first that g is orientation-preserving. There is, by the Mean Value Theorem, a point in each component of $Y \setminus \{p\}$ with derivative $\geq 1 + \kappa$. By the Minimum Principle, the derivative at p is $\geq 1 + \kappa$. If g is orientation-reversing, apply this argument to g^2 . \square

Lemma 10.6. *Given $\kappa > 0$, there are $\gamma, \alpha > 0$ such that the following holds. Let $Z \subset Y \subset g(Z)$ be intervals, where g is a C^2 diffeomorphism from Y to $g(Y)$ with non-positive Schwarzian derivative. Suppose*

$$\text{dist}(g(Z), \partial g(Y)) \geq \kappa |g(Z)|.$$

If $|Z| \leq \gamma |g(Z)|$ then $|Dg|_Z| > 2$. If $|Z| \geq \gamma |g(Z)|$ then $g^{-2}(Z)$ and the components of $Z \setminus g^{-2}(Z)$ all have length bounded by $|Z|(1 - \alpha)$.

Proof. The first statement is obvious, by bounded distortion. For the second statement, since $Z \geq \gamma |g(Z)|$, by bounded distortion there is a uniform lower bound on $|g^{-2}(Z)|/|Z|$, giving the upper bound on the components of $Z \setminus g^{-2}(Z)$. By Lemma 10.5, at the fixed point p of g , $Dg^2 \geq 1 + \kappa$, so there is a uniform neighbourhood of p with $Dg > \sqrt{1 + \kappa}$. Using the Minimum Principle, a uniform upper bound on $|g^{-2}(Z)|$ follows easily. \square

Lemma 10.7. *Given $\kappa > 0$, there is an $N \geq 1$ such that the following holds. Let (\hat{X}, \hat{F}, τ) be a level- R induced map for a d -branched piecewise-monotone map f with non-positive Schwarzian derivative. Suppose $\kappa \leq \min\{|Y| : Y \in \hat{\mathcal{P}}_R, Y \subset \hat{I}^R\}$. Then $|D\hat{F}^{NK}| > 2$, where $K \geq 1$ depends only on d and R .*

Proof. For each domain D of \hat{I}^R containing a point of \hat{X} , we have $\text{dist}(\hat{X}, \partial D) > \kappa$, while $|D| \leq 1 = |I|$. Let $n \geq 1$ and let $Z \subset \hat{X}$ be a branch of \hat{F}^n , so \hat{F}^n maps Z homeomorphically onto a connected component of \hat{X} contained in a domain D of \hat{I}^R , with $\hat{F}^n|_Z = \hat{f}^j$ say. There is, by Lemma 4.2, an interval $Y \supset Z$ mapped homeomorphically by \hat{f}^j onto D . Since $j \geq n$, $Y \subset \hat{X}$. Bounded distortion then gives a positive lower bound, depending only on κ , for $|D\hat{F}^n|$.

Let \hat{G} denote the first return map to a connected component \hat{X}_G of \hat{X} . Since there are at most a finite number (depending on d, R) of connected components, it suffices to show that there is an N for which $|D\hat{G}^N| > 2$. The general result follows because high iterates can be decomposed into a sequence of repeated first return maps bookended by two maps with derivative bounded below.

So let Z be a branch of \hat{G}^n , for some $n \geq 1$, with $\hat{G}^n|_Z = \hat{f}^j$ say. As before, there is an interval $Y \supset Z$ mapped homeomorphically by \hat{f}^j onto D the domain of \hat{I}^R , now containing Z .

Let $\gamma, \alpha > 0$ be given by Lemma 10.6. If $|Z| \leq \gamma |\hat{X}_G|$ then $|D\hat{G}^n| > 2$. Otherwise, by the second part of Lemma 10.6, the branches of \hat{G}^{3n} contained in Z have length bounded by $|Z|(1 - \alpha)$. An inductive argument gives some N for which all branches of \hat{G}^N have length at most $\gamma |\hat{X}_G|$, and thus $|D\hat{G}^N| > 2$, the required derivative estimate. \square

Denote by $\hat{\mathcal{K}}$ the set of periodic points $\hat{p} \in \hat{I}$ with multiplier 1, that is, for which, for some $n \geq 1$, $\hat{f}^n(\hat{p}) = \hat{p}$ and $D\hat{f}^n(\hat{p}) = 1$. It follows from the Minimum Principle that if $\hat{p} \in D \in \mathcal{D}$, then on at least one component of $D \setminus \{\hat{p}\}$, \hat{f}^n is the identity and the component lies in $\hat{\mathcal{K}}$.

Lemma 10.8. *Every recurrent point in $\hat{I} \setminus \hat{\mathcal{K}}$ lies in the domain of a level- R induced map for all sufficiently large R .*

Proof. Suppose $\hat{x} \in \hat{I} \setminus \hat{\mathcal{K}}$ is recurrent. Then there is a sequence of cylinder sets $\hat{Z}_k \ni \hat{x}$ with $\hat{f}^{n_k}(\hat{Z}_k) = D$, where D is the domain of \hat{I} containing \hat{x} . From the Minimum Principle, it follows that $|\hat{Z}_k| \rightarrow 0$ as $n_k \rightarrow \infty$. Points in the boundary of cylinder sets are not recurrent. Therefore \hat{x} lies in a non-boundary R -cylinder for large R . \square

We remark that we do not have exponential growth rates (necessarily), but at least $\limsup_{n \rightarrow \infty} |D\hat{f}^n(\hat{x})| = \infty$ for recurrent points $\hat{x} \notin \hat{\mathcal{K}}$. By Lemmas 10.7-10.8 and Birkhoff's theorem, one obtains:

Corollary 10.9. *Every measure $\hat{\mu} \in \mathcal{M}_{\hat{f}}$, not supported on $\hat{\mathcal{K}}$, has positive Lyapunov exponent.*

Given (\hat{X}, \hat{F}, τ) or its embedding (X, F, τ) , we say that $\log |D\hat{F}|$ (or equivalently $\log |DF|$) is *locally Hölder* if there are $C_0, \alpha > 0$ such that, for each n and on every branch Z of F^n ,

$$|\log |DF(x)| - \log |DF(y)|| \leq C_0 \exp(-\alpha n).$$

Lemma 10.10. *Let $f \in \mathcal{F} \in \mathcal{F}_M$ and let (\hat{X}, \hat{F}, τ) be the level- R induced map. Then $\log |DF|$ is locally Hölder.*

Proof. For ease of notation, we drop the hats and work with the embedded induced map F . We have

$$\kappa := \min\{|Y| : Y \in \hat{P}_R, Y \subset \hat{I}^R\} > 0.$$

By Lemma 10.7, there is some N_0 for which $|DF^{N_0}| > 2$. The length of the connected components of X are bounded by $1 = |I|$. If Z is a branch of F^n with $n \geq kN_0 + 1$ say, then $|F(Z)| < \exp(-2k)$, since $2 < |DF^{N_0}|$. The map F on Z extends with non-positive Schwarzian derivative to map a larger interval homeomorphically onto an interval X' with

$$\text{dist}(F(Z), \partial X') \geq \kappa \geq \kappa |F(Z)| \exp(2k).$$

Applying the Koebe Lemma 10.2 gives $|\log |DF(x)| - \log |DF(y)|| \leq C_0 \exp(-2k)$ for all $x, y \in Z$, for some constant C_0 depending just on κ . Now replacing k by $(n-1)/N_0$, we obtain

$$|\log |DF(x)| - \log |DF(y)|| \leq C_1 \exp(-n/N_0),$$

for some constant C_1 , as required. \square

11. ALMOST UPPER-SEMICONTINUITY OF THE FREE ENERGY

In this section we prove Theorem 1.18.

So, let $(f_k)_k$ be a sequence of d -branched maps with non-positive Schwarzian derivative and converging to f_0 as $k \rightarrow \infty$ with decreasing critical relations. Let $\varepsilon' > 0$ and let $\mu_k \in \mathcal{M}_{f_k}$ for $k \geq 1$, with $h(\mu_k) \geq \varepsilon'$. Let t_k converge to $t_0 \in \mathbb{R}$ and assume $E(f_k, \mu_k, -t_k \log |Df_k|)$ converges to a limit E_L . Let $R_0 \geq 8d$ be big enough that $2\varepsilon(R_0) < \varepsilon'$, where $\varepsilon(R_0)$ is defined in (2.1). For each $R \geq R_0$, we could construct in §6 sets $\Omega_k, \Omega_k^N, \Omega'_k$. Taking subsequences, we can assume that

- μ_k converge to a limit measure denoted μ_∞ ;
- for each $R \geq R_0$ and $N \geq 1$, the sequences $(\Omega_k^N)_k$ and $(\Omega'_k)_k$ are eventually constant.

This allows us, for each R , to obtain convergent sequences of embedded level- R induced maps

$$(X^k(R) = X^k, F_k, \tau_k) \rightarrow (Y^0, F_0, \tau_0).$$

For $k \geq 1$, we obtain the embedded induced measures $\nu_k^R \in \mathcal{M}_{F_k}$ via (8.2). Note we do not indicate through notation the dependence of F_k on R , but this should be understood. Again taking subsequences, we assume that for each R , as $k \rightarrow \infty$,

- ν_k^R converges to a limit measure ν_0^R ;
- T_k^R , the integral of the inducing time, converges to some T_L^R ;

- $h(\nu_k^R)$ converges to a limit h_L^R ;
- $\lambda(\nu_k^R) := \int \log |DF_k| d\nu_k^R$ converges to some limit λ_L^R .

By Lemma 9.1, $\nu_0^R \in \mathcal{M}_{F_0}$. Set $T_\Delta^R := T_L^R - T_0^R \geq 0$ (by (9.4)). Note that $1 \leq T_k^R \leq 1/\eta(R_0, d)$ for each $k \geq 0, R \geq R_0$, where η is given by Lemma 8.2. As $T_0^R \geq 1$,

$$T_\Delta^R \leq 1/\eta(R_0, d). \quad (11.1)$$

We can spread the measure $\nu_0 = \nu_0^R$ to get a f_0 -invariant probability measure μ_0^R , defined by

$$\mu_0^R = \frac{1}{T_0} \sum_n \sum_{j=0}^{n-1} f_{0*}^j(\pi_{1*} \nu_0|_{\{\tau_0=n\}}).$$

Since

$$\nu_0^R = \lim_{k \rightarrow \infty} \nu_k^R \ll \lim_{k \rightarrow \infty} \mu_k,$$

μ_0^R is a light limit measure (see Definition 1.2) for the sequence $(\mu_k)_k$. Upon taking further subsequences, we similarly obtain measures $\mu_0^{R'}$ for each $R' > R$, and we can assume furthermore that $E(\mu_k)$ converges to a limit E_L .

Set $\lambda_\Delta^R := \lambda_L^R - \lambda_0^R$. Due to non-positive Schwarzian derivative, Lemma 10.3 gives us a uniform lower bound on $|DF_k|$. There is a uniform upper bound on $|Df_k|$, so $\log |DF_k| \leq \tau_k C$ for some uniform constant $C > 1$ (independent of R). Hence

$$0 \leq \lambda_\Delta^R = \lim_{K \rightarrow \infty} \lim_{k \rightarrow \infty} \int_{\{\tau_k \geq K\}} \log |DF_k| d\nu_k^R \leq T_\Delta^R \log C. \quad (11.2)$$

We shall use (8.4) and (8.5) repeatedly in the following lemmas.

Lemma 11.1. *If $\lim_{R \rightarrow \infty} T_\Delta^R = 0$, then μ_0^R converges strongly to μ_∞ . Moreover, $\lambda(\mu_\infty) = \lim_{k \rightarrow \infty} \lambda(\mu_k)$ and*

$$E(\mu_\infty) \geq E_L.$$

Proof. The first statement follows immediately from Lemma 9.2. From (11.2), if $\lim_{R \rightarrow \infty} T_\Delta^R = 0$ then $\lim_{R \rightarrow \infty} \lambda_\Delta^R = 0$ and

$$\begin{aligned} \int \log |Df_0| d\mu_\infty &= \lim_{R \rightarrow \infty} \int \log |Df_0| d\mu_0^R \\ &= \lim_{R \rightarrow \infty} \frac{\lambda_0^R}{T_0^R} \\ &= \lim_{R \rightarrow \infty} \lim_{k \rightarrow \infty} \frac{\lambda_k^R}{T_L^R} \\ &= \lim_{k \rightarrow \infty} \int \log |Df_k| d\mu_k. \end{aligned}$$

Thus $E(\mu_\infty) - E_L = h(\mu_\infty) - \lim_{k \rightarrow \infty} h(\mu_k)$. By Theorem 1.15, entropy is upper-semicontinuous, we deduce that $E(\mu_\infty) \geq E_L$. \square

If $T_\Delta^R \neq 0$ then any jump in free energy can be written in a suggestive form.

Lemma 11.2. *Suppose $T_\Delta^R \neq 0$. Then*

$$\frac{T_0}{T_\Delta} (E_L - E(\mu_0)) = \frac{h_\Delta - t\lambda_\Delta}{T_\Delta} - E_L.$$

Proof. From

$$E_L = \lim_{k \rightarrow \infty} \frac{h(\nu_k) - t\lambda(\nu_k)}{T_k} = \frac{h_L - t\lambda_L}{T_L},$$

one deduces (to verify, substitute $T_L - T_0$ for T_Δ) that

$$E_L = \frac{h_L - t\lambda_L - T_\Delta E_L}{T_0}.$$

Consequently, knowing $E(\mu_0) = \frac{h(\nu_0) - t\lambda(\nu_0)}{T_0}$,

$$E_L - E(\mu_0) = \frac{1}{T_0} ((h_\Delta - t\lambda_\Delta) - T_\Delta E_L). \quad (11.3)$$

Multiply both sides by T_0/T_Δ to conclude. \square

Let $T_\Delta^\infty := \liminf_{R \rightarrow \infty} T_\Delta^R \geq 0$.

Lemma 11.3. *If $T_\Delta^\infty > 0$ then there exists $R_0, \theta > 0$ such that for all $R \geq R_0$, the light limit measure μ_0^R has entropy $h(\mu_0^R) \geq \lim_{k \rightarrow \infty} h(\mu_k) + \theta$.*

Proof. For all R large enough, $T_\Delta^R > T_\Delta^\infty/2$ and

$$h(\mu_k) \frac{T_\Delta^\infty}{2T_0^R} - 6\varepsilon(R) \frac{T_k^R}{T_0^R} > \theta$$

for all large k and some $\theta > 0$.

Then for all large k ,

$$\begin{aligned} h(\mu_0^R) &= \frac{h(\nu_0^R)}{T_0^R} \\ &\geq \frac{h(\nu_k) - 5\varepsilon(R)T_\Delta^R}{T_0^R} \\ &= h(\mu_k) \frac{T_k^R}{T_0^R} - 5\varepsilon(R) \frac{T_\Delta^R}{T_0^R} \\ &\geq h(\mu_k) + h(\mu_k) \frac{T_k^R - T_0^R}{T_0^R} - 5\varepsilon(R) \frac{T_\Delta^R}{T_0^R} \\ &> h(\mu_k) + h(\mu_k) \frac{T_\Delta^\infty}{2T_0^R} - 6\varepsilon(R) \frac{T_\Delta^\infty}{T_0^R} \\ &\geq h(\mu_k) + \theta, \end{aligned}$$

completing the proof. \square

Lemma 11.4. *Suppose $T_\Delta^\infty > 0$ and let R_0 be given by Lemma 11.3. If $E(\mu_0^{R_0}) < E_L$, either there is a light limit measure with free energy strictly greater than E_L or there is a $\gamma > 0$ such that $E(\mu_0^R) \leq E_L - \gamma$ for all $R \geq R_0$.*

Proof. Suppose $E(\mu_0^{R_0}) = E_L - \gamma'$ for some $\gamma' > 0$. Let $R \geq R_0$ and consider the measures ν_k^R . Note that $\hat{\nu}_k^{R_0}, \hat{\nu}_k^R$ are constructed in the Hofbauer extension as normalised restrictions of $\hat{\mu}_k$. Thus for subsets of $X(R_0)$,

$$\nu_k^R = \frac{T_k^R}{T_k^{R_0}} \nu_k^{R_0},$$

so on subsets of Y_0 ,

$$\nu_0^R = \frac{T_L^R}{T_L^{R_0}} \nu_0^{R_0}.$$

Thus we can write $\mu_0^R = (1/T_L^{R_0})\mu_0^{R_0} + (1 - 1/T_L^{R_0})\mu'$, say, where μ' is a positive f_0 -invariant probability measure (not necessarily different from $\mu_0^{R_0}$). In particular, μ^R and μ' are light limit measures. Either $E(\mu') > E_L$ or, since E is linear,

$$\begin{aligned} E(\mu_0^R) &\leq \frac{E_L - \gamma'}{T_L^{R_0}} + \left(1 - \frac{1}{T_L^{R_0}}\right) E_L \\ &\leq E_L - \gamma'/T_L^{R_0}. \end{aligned}$$

Setting $\gamma := \gamma'/T_L^{R_0}$ completes the proof. \square

Lemma 11.5. *Suppose $E(\mu_0^R) \leq E_L - \gamma$ for all $R \geq R_0$, for some $R_0, \gamma > 0$. Then for sufficiently large R and k , there exists $\mu \in \mathcal{M}_{f_k}$ supported on a periodic orbit, with $E(\mu) > E_L + \eta(R_0, d)\gamma/3$.*

Proof. From Lemma 11.1 it follows that $T_\Delta^\infty > 0$. Let us fix $R \geq R_0$ large enough that $\varepsilon := \varepsilon(R) < \gamma\eta(R_0, d)/10$. For this R , we have $X^k, Y^0, F_k, \tau_k, \nu_k, T_k$ for each k . We write $T_L, T_\Delta, h_L, \lambda_L$ for $T_L^R, T_\Delta^R, h_L^R, \lambda_L^R$. Set $h_\Delta := h_L - h(\nu_0)$ and $\lambda_\Delta := \lambda_L - \lambda(\nu_0)$. Let μ_0 denote the projection of ν_0 .

By Lemma 11.2,

$$E_L - E(\mu_0) = \frac{1}{T_0} ((h_\Delta - t\lambda_\Delta) - T_\Delta E_L) \geq \gamma.$$

Noting $T_0 \geq 1$,

$$-t\lambda_\Delta/T_\Delta \geq E_L + \gamma/T_\Delta - h_\Delta/T_\Delta.$$

By (9.1), $1/T_\Delta \geq \eta(R_0, d)$. By Lemma 9.3, $h_\Delta \leq 5\varepsilon T_\Delta$. These estimates, combined with choice of R , imply

$$-t\lambda_\Delta/T_\Delta \geq E_L + \eta(R_0, d)\gamma/2. \quad (11.4)$$

As previously, let ξ_k denote the collection of branches of F_k . Let

$$\xi_k^N := \bigvee_{j=0}^{N-1} F_k^{-j} \xi_k.$$

Let $\xi_k(K)$ denote the elements of ξ_k on which $\tau_k \geq K$ and set

$$V_K^k := \bigcup_{Z \in \xi_k(K)} Z.$$

Let $x \in V_K^k$ be a typical point for the measure ν_k (we know T_Δ^R is positive, so V_K^k has positive measure), so x is recurrent. Now we can speed up the map F_k a little, and define G , the first return map (for F_k) to V_K^k . Let n, w be given by Lemma 7.2, so for some arbitrarily large n , w is an n -periodic point (for F_k) and $G^j(w), G^j(x)$ lie in the same element of $\xi_k(K)$ for $j = 0, \dots, n-1$. The points w will still be periodic for G . We wish to estimate the Lyapunov exponent of w . Let $e(n)$ denote the n^{th} return time of x to V_K^k under F , so $F^{e(j)}(x)$ and $G^j(w)$ are in the same element of $\xi_k(K)$, so $\tau_k(G^j(w)) = \tau_k(F^{e(j)}(x))$. By Birkhoff's theorem,

$$\frac{n}{e(n)} \approx \nu_k(V_K^k) \leq (T_L + 1)/K,$$

which tends to 0 as $K \rightarrow \infty$. Let

$$g(y) := \sum_{i=0}^{e_1(y)-1} \tau_k(F_k^i(y)).$$

If $\tau_k(F_k(G^j(w))) \geq K$, then τ_k and g coincide at $G^j(w)$. If $\tau_k(F_k(G^j(w))) < K$, then $g(G^j(w)) = \tau_k(G^j(w)) + g(F_k(G^j(w)))$. Consequently, by Lemma 7.2,

$$K \leq \tau_k(G^j(w)) \leq g(G^j(w)) \leq \tau_k(G^j(w)) + M.$$

We then obtain the following estimates, using χ to denote the characteristic function of a set. Recall that $\lim_{K \rightarrow \infty} \lim_{k \rightarrow \infty} \int_{V_K^k} \tau_k d\nu_k = T_\Delta$. In what follows, $\kappa(K, k, n)$ may vary from line to line, but it will satisfy

$$\lim_{K \rightarrow \infty} \lim_{k \rightarrow \infty} \lim_{n \rightarrow \infty} \kappa(K, k, n) = 0.$$

$$\begin{aligned} \sum_{j=0}^{n-1} g(G^j(w)) &= (1 + \kappa(K, k, n)) \sum_{j=0}^{n-1} \tau_k(G^j(w)) \\ &= (1 + \kappa(K, k, n)) \sum_{j=0}^{e(n)-1} \chi_{V_K^k}(F_k^j(x)) \tau_k(F_k^j(x)) \\ &= e(n) \left(\kappa(K, k, n) + \int_{V_K^k} \tau_k d\nu_k \right) \\ &= e(n) (\kappa(K, k, n) + T_\Delta). \end{aligned}$$

To estimate the derivative, note that distortion is uniformly bounded and that the derivative corresponding to quick words is uniformly bounded below away from zero (by Lemma 10.3), while it is bounded above by $\sup_k \sup |Df_k|^M$. Thus

$$\begin{aligned} \log |DG^n(w)| &= (1 + \kappa(K, k, n)) \sum_{j=0}^{e(n)-1} \chi_{V_K^k}(F_k^j(x)) \log |DF_k(F_k^j(x))| \\ &= e(n) \left(\kappa(K, k, n) + \int_{V_K^k} \log |DF_k| d\nu_k \right) \\ &= e(n) (\kappa(K, k, n) + \lambda_\Delta). \end{aligned}$$

Dividing, the Lyapunov exponent for the f_k -invariant equidistribution, denoted μ_w say, along the periodic orbit of $\pi_1 w \in I$ is $\lambda_\Delta/T_\Delta + \kappa(K, k, n)$. The free energy of the equidistribution satisfies, using (11.4) and (9.11),

$$\begin{aligned} E(\mu_w) &= -t\lambda_\Delta/T_\Delta + \kappa(K, k, n) \\ &\geq E_L + \eta(R_0, d)\gamma/2 + \kappa(K, k, n) \\ &> E_L + \eta(R_0, d)\gamma/3. \end{aligned}$$

Thus for each k large enough, there is a f_k -invariant measure $\mu^k (= \mu_w, \text{ for some } K, k, n)$, supported on a periodic orbit, with $E(f_k, \mu^k) > E_L + \eta(R_0, d)\gamma/3$. This completes the proof of Lemma 11.5. \square

11.1. Proof of Theorem 1.18.

Proof. In the case $T_\Delta^\infty = 0$, just apply Lemma 11.1: the second alternative of Theorem 1.18 holds. If $T_\Delta^\infty > 0$, by Lemma 11.3, the light limit measure $\mu_0^{R_0}$ has entropy strictly greater than $\inf_k h(\mu_k)$. If this measure has free energy at least E_L , the third alternative of the theorem holds, or if some other light limit measure has free energy $> E_L$ the first alternative of the theorem holds. If these alternatives do not hold, Lemma 11.4 gives a $\gamma > 0$ for which $E(\mu_0^R) \leq E_L - \gamma$ for all $R \geq R_0$. We then derive a contradiction via Lemma 11.5 and the hypothesis $\liminf_{k \rightarrow \infty} E_+(f_k) \leq E_L$. \square

12. KATOK THEORY

An idea due, we believe, to Katok is that one can approximate measures with non-zero Lyapunov exponents by measures on hyperbolic sets with similar entropy and exponents, see [PU, §11.6] for this in the complex dynamics setting.

Lemma 12.1. *Given $f \in \mathcal{F} \in \mathcal{F}_{\text{NSD}}$, suppose $\mu \in \mathcal{M}_f$ with positive entropy and (therefore) positive Lyapunov exponent. Then there exists a sequence $(\Lambda_n)_{n \geq 1}$ with the follow properties. Each Λ_n is a forward-invariant, compact, hyperbolic subset of I containing periodic points. If $(\mu_n)_{n \geq 1}$ is any sequence of measures $\mu_n \in \mathcal{M}_f$ with support Λ_n , then as $n \rightarrow \infty$,*

$$\begin{aligned} \mu_n &\rightarrow \mu, \\ \lambda(\mu_n) &\rightarrow \lambda(\mu). \end{aligned}$$

Moreover, if one takes as μ_n the measure of maximal entropy for $f|_{\Lambda_n}$, then $h(\mu_n) \rightarrow h(\mu)$.

Proof. To show convergence of measures, let ψ_1, ψ_2, \dots be a sequence of Lipschitz continuous real-valued functions on I . It suffices to show that, given $N \geq 1$ and $\alpha > 0$, Birkhoff averages of ψ_i along orbits in Λ_n approximate $\int \psi_i d\mu$ to within α for all $i \leq N$ and all large n .

For some $R \geq 1$, there exists, in the Hofbauer extension, a non-boundary element \hat{Z} of $\hat{\mathcal{P}}_R$ with $\theta := \hat{\mu}(\hat{Z}) > 0$.

Let \hat{G} denote the first return map to \hat{Z} with n^{th} return time e_n . By Lemma 10.7, $|D\hat{G}^n| > 2$ for all large n . The map \hat{G} has inducing time τ with $\int_{\hat{Z}} \tau d\hat{\mu} = 1$, while the probability measure $\hat{\nu} := \theta^{-1}\hat{\mu}$ is \hat{G} -invariant with entropy $h(\hat{\nu}) = h(\mu)/\theta$ (recall (8.5)).

Let $\mathcal{Q} = \mathcal{Q}_1$ denote the branch partition of \hat{G} , that is, the collection of connected components of the domain of \hat{G} . It is generating and has entropy equal to $h(\hat{\nu})$. Denote by \mathcal{Q}_n the join $\bigvee_{j=1}^n \hat{G}^{-j+1}\mathcal{Q}_1$ and by $\mathcal{Q}_n[\hat{x}]$ the element of \mathcal{Q}_n containing \hat{x} .

For almost every \hat{x} in \hat{Z} , as $n \rightarrow \infty$, by Shannon-MacMillan-Breiman and Birkhoff's theorems, for $1 \leq i \leq N$,

$$\begin{aligned} \frac{-1}{n} \log \hat{\nu}(\mathcal{Q}_n[\hat{x}]) &\rightarrow h(\hat{\nu}), \\ e_n(\hat{x})/n &\rightarrow \theta^{-1}, \\ \frac{1}{n} \log |D\hat{G}^n(\hat{x})| &\rightarrow \lambda(\mu)/\theta, \\ \lim_{k \rightarrow \infty} \frac{1}{k} \sum_{j=0}^{k-1} \psi_i \circ \pi \circ \hat{f}^j(\hat{x}) &= \int \psi_i d\mu. \end{aligned}$$

Since a set of measure α , say, cannot be covered by fewer than $\alpha\varepsilon^{-1}$ sets of measure $\leq \varepsilon$, we obtain the following, using little- o notation.

There is a finite collection \mathcal{Q}'_n of $\exp(nh(\hat{\nu})(1+o(1)))$ elements of \mathcal{Q}_n on which

$$e_n = \frac{n}{\theta}(1+o(1)), \quad (12.1)$$

$$|D\hat{G}^n| = \exp\left(\frac{n}{\theta}\lambda(\mu)(1+o(1))\right), \quad (12.2)$$

$$\sum_{j=0}^{e_n-1} \psi_i \circ \pi \circ \hat{f}^j = (1+o(1)) \int \psi_i d\mu. \quad (12.3)$$

Note that to obtain uniformity, (12.3) relies on some iterate of G being expanding and on Lipschitz continuity of the functions ψ_i .

Let $\hat{Y}_n := \cup_{\hat{V} \in \mathcal{Q}'_n} \hat{V}$. Set

$$\hat{\Lambda}'_n := \bigcap_{k \geq 0} \overline{\hat{G}^{-kn}(\hat{Y}_n)} = \bigcap_{k \geq 0} \hat{G}^{-kn}(\hat{Y}_n),$$

the latter equality holding because $\overline{\hat{Y}_n} \subset \hat{Z}$ and \hat{G}^n maps components of \hat{Y}_n onto \hat{Z} . Then all iterates of \hat{G}^n are defined on $\hat{\Lambda}'_n$ and $\hat{\Lambda}'_n$ is a forward-invariant set under \hat{G}_n , while e_n is bounded. Since we only selected a finite number of branches, $\hat{\Lambda}'_n$ is compact. Thus

$$\Lambda_n := \bigcup_{k \geq 0} f^k(\pi(\Lambda'_n))$$

is a forward-invariant, compact hyperbolic subset of I (also known as a *conformal expanding repeller*). The fixed points of \hat{G}^n in $\hat{\Lambda}'_n$ correspond to periodic points for f in Λ_n .

Every point in Λ_n enters $\pi(\Lambda'_n)$ infinitely often. Thus for all $x \in \Lambda_n$, by (12.2)-(12.3), for $k \geq n^2$, say,

$$\begin{aligned} \frac{1}{k} \log |Df^k(x)| &= (1+o(1))\lambda(\mu), \\ \sum_{j=0}^{k-1} \psi_i \circ f^j(x) &= (1+o(1)) \int \psi_i d\mu. \end{aligned}$$

In particular, any measure μ_n supported on Λ_n has Lyapunov exponent

$$\lambda(\mu)(1+o(1)),$$

and for any sequence $(\Lambda_n, \mu_n)_n$, μ_n converges to μ .

The topological entropy of $\hat{G}^n : \hat{Y}_n \rightarrow \hat{Z}$ is

$$\log \#Q'_n = nh(\hat{\nu})(1 + o(1)).$$

The average inducing time for any measure ν_n on $\hat{\Lambda}'_n$ is

$$T_n := \int_{\hat{Y}_n} e_n d\nu_n = \frac{n}{\theta}(1 + o(1)).$$

Consequently (use (8.4) in both directions), the measure of maximal entropy for f restricted to Λ_n has entropy

$$nh(\hat{\nu})(1 + o(1))/T_n = \theta h(\hat{\nu})(1 + o(1)) = h(\mu)(1 + o(1))$$

as required. \square

Corollary 12.2. *Suppose $f \in \mathcal{F} \in \mathcal{F}_{\text{NSD}}$, $\varepsilon > 0$, and $\mu \in \mathcal{M}_f$ has positive entropy. For each g sufficiently close to f in \mathcal{F} , there is a measure $\mu_g \in \mathcal{M}_g$ with*

$$|h(\mu_g) - h(\mu)| < \varepsilon$$

and

$$|\lambda(\mu_g) - \lambda(\mu)| < \varepsilon.$$

In particular, $\liminf_{g \rightarrow f} P(g, -t \log |Dg|) \geq E(f, \mu, -t \log |Df|)$.

Proof. By Lemma 12.1, there is a measure $\nu \in \mathcal{M}_f$ supported on a compact hyperbolic set Λ with $|h(\nu) - h(\mu)| < \varepsilon/2$ and $|\lambda(\nu) - \lambda(\mu)| < \varepsilon/2$. For g sufficiently C^1 -close to f_0 (on a neighbourhood of Λ), there is a smooth conjugacy h_g (close to the identity) on a neighbourhood of Λ with $g = h \circ f \circ h^{-1}$, so $h_*\nu \in \mathcal{M}_g$, and the entropy and Lyapunov exponent of $h_*\nu$ approximate those of ν to within $\varepsilon/2$. The result follows. \square

13. ANCILLARY PROOFS

Proof of Corollary 1.20. For $t \in (t^-, t^+)$, consider a sequence of measures $\mu_k \in \mathcal{M}_f$ with $E(\mu_k) \rightarrow P(-t \log |Df|)$. By hypothesis, for large k ,

$$h(\mu_k) - t \int \log |Df| d\mu_k \geq \frac{P(-t \log |Df|) - P^0(-t \log |Df|)}{2} + P^0(-t \log |Df|).$$

In Lemma 12.1, it is shown that positive entropy measures can be approximated by equidistributions on periodic orbits with similar Lyapunov exponents. Hence

$$P^0(-t \log |Df|) - \left(-t \int \log |Df| d\mu \right) \geq 0$$

for every $\mu \in \mathcal{M}_f$ (not just those with zero entropy, as in the definition). Therefore,

$$h(\mu_k) \geq \frac{P(-t \log |Df|) - P^0(-t \log |Df|)}{2} > 0,$$

so the measures have uniformly positive entropy. Applying Theorem 1.18, we obtain existence of an equilibrium measure with positive entropy. \square

Proof of Proposition 1.21. Let

$$\underline{\lambda} := \inf_{\mu \in \mathcal{M}_f} \lambda(\mu) = \inf_{\mu \in \mathcal{M}_f, h(\mu)=0} \lambda(\mu),$$

the equality holding by Lemma 12.1. For $t \geq 0$, $P^0(-t \log |Df|) = -t\underline{\lambda}$. Suppose μ is an equilibrium state for $-t^+ \log |Df|$ with $h(\mu) > 0$. Then the lines $t \mapsto -t\underline{\lambda}$ and $t \mapsto h(\mu) - t\lambda(\mu)$

intersect at $(t^+, P(-t^+ \log |Df|))$ and subtend the graph of $t \mapsto P(-t \log |Df|)$, showing lack of differentiability. In the other direction, take a strictly increasing sequence of t_k converging to t^+ and let μ_k be the corresponding equilibrium states. Non-differentiability together with convexity implies that for some $\delta > 0$, $\lambda(\mu_k) \geq \lambda_{\min} + \delta$ for all k . Simple geometry then implies that $\inf_k h(\mu_k) > 0$ and that $E(f, \mu_k, -t^+ \log |Df|) \rightarrow P(-t^+ \log |Df|)$. Applying Theorem 1.18 gives existence of the equilibrium state with positive entropy. The proof for t^- is the same (with $t_k \searrow t^-$ and $\bar{\lambda}$, appropriately defined, in place of $\underline{\lambda}$). \square

Lemma 13.1. *Let $\varepsilon > 0$ and $t \in \mathbb{R}$. Let $(f_k)_k \in \mathcal{F}_{\text{NSD}}$ be a sequence converging to f_0 as $k \rightarrow \infty$ with decreasing critical relations. Suppose that*

- (a) *for each $k \geq 0$ there is an equilibrium state μ_k for the map f_k and the potential $-t \log |Df_k|$;*
- (b) *each μ_k has entropy at least ε .*

Then

$$\lim_{k \rightarrow \infty} P(f_k, -t \log |Df_k|) = P(f_0, -t \log |Df_0|). \quad (13.1)$$

Proof. From Corollary 12.2,

$$\liminf_{k \rightarrow \infty} P(f_k, -t \log |Df_k|) \geq P(f_0, -t \log |Df_0|).$$

By Theorem 1.18, there is a positive entropy light limit measure $\mu'_0 \in \mathcal{M}_{f_0}$ with

$$E(\mu'_0) \geq \limsup_{k \rightarrow \infty} P_{f_k}(-t \log |Df_k|).$$

By definition of pressure,

$$P(f_0, -t \log |Df_0|) \geq E(\mu'_0),$$

so the limit of $P(f_k, -t \log |Df_k|)$ exists and coincides with $P(f_0, -t \log |Df_0|)$. \square

Lemma 13.2. *Let $(f_k)_{k \geq 0} \in \mathcal{F}_{\text{NSD}}$ be a sequence converging to f_0 as $k \rightarrow \infty$ with decreasing critical relations. Let $t_0 \in \mathbb{R}$. Suppose that*

- (a) *$\mu_k \rightarrow \mu_\infty$ as $k \rightarrow \infty$, where $\mu_k \in \mathcal{M}_{f_k}$ for $k \geq 1$;*
- (b) *$E_L := \lim_{k \rightarrow \infty} E(f_k, \mu_k, -t_0 \log |Df_k|) = P(-t_0 \log |Df_0|)$;*
- (c) *$E_L \geq \liminf E_+(f_k, -t_0 \log |Df_k|)$;*
- (d) *$\liminf_{k \rightarrow \infty} h(\mu_k) > 0$;*
- (e) *the pressure functions $t \mapsto P(f_k, -t \log |Df_k|)$ converge on a neighbourhood of t_0 to a limit function $t \mapsto p(t)$.*

If μ_∞ is not a convex combination of equilibrium states for the potential $-t_0 \log |Df_0|$, then some light limit measure μ_ is an equilibrium state with*

$$h(\mu_*) > \liminf_{k \rightarrow \infty} h(\mu_k),$$

and p is not differentiable at t_0 .

Proof. Considering the free energies of μ_k , the graph of p is subtended by the line passing through the points $(0, \liminf_{k \rightarrow \infty} h(\mu_k))$, $(t_0, p(t_0))$.

One of the last two alternatives of Theorem 1.18 hold. If μ_∞ is not a convex combination of equilibrium states, the second alternative does not hold and the third (final) alternative must hold. Thus there is a light limit measure μ_* with $E(\mu_*) = P(f, -t_0, \log |Df_0|)$ and with entropy $h(\mu_*) > \liminf_{k \rightarrow \infty} h(\mu_k)$. Via Corollary 12.2, considering free energies of the given measures

for maps close to f_0 , one obtains that the graph of p is subtended by the line passing through the points $(0, h(\mu_*)), (t_0, p(t_0))$.

As a result, p is not differentiable at t_0 . \square

Proof of Theorem 1.22. By hypothesis,

$$P(f_k, -t_0 \log |Df_k|) \geq P^0(f_k, -t_0 \log |Df_k|) + \varepsilon$$

for each k . Since the pressure functions are uniformly Lipschitz, we immediately deduce the existence of a neighbourhood U of t_0 for which, for all large k ,

$$P(f_k, -t \log |Df_k|) \geq P^0(f_k, -t \log |Df_k|) + \varepsilon/2 \quad (13.2)$$

for all $t \in U$. By Corollary 1.20, there is an equilibrium state $\mu_{k,t}$ for f_k and the potential $-t \log |Df_k|$ (for large k and for $k = 0$). By (13.2), $h(\mu_{k,t}) \geq \varepsilon/2$. Applying Lemma 13.1, the pressures converge for $t \in U$:

$$\lim_{k \rightarrow \infty} P(f_k, -t \log |Df_k|) = P(f_0, -t \log |Df_0|).$$

Since $t \mapsto P(f_0, -t \log |Df_0|)$ is assumed differentiable at t_0 , Lemma 13.2 implies that any limit measure of the sequence $(\mu_k)_k$ is a convex combination of equilibrium measures. \square

14. THERMODYNAMICS FOR INDUCED MAPS

A strategy used to study interval maps which lack Markov structure and expansiveness is to consider induced maps, a generalisation of the first return map. We shall use level- R primitive induced maps $(\mathcal{Y}, \hat{F}_{\mathcal{Y}}, \tau)$ as introduced in Definition 4.16. These maps will be expanding and Markovian, although over a countable alphabet and not full-branched. The idea is to study (thermodynamic) properties of the induced maps and then to translate the results into the original system. We shall make use of thermodynamic formalism for topological Markov shifts (in Sarig's terminology) as developed by Sarig and by Mauldin and Urbański. We then prove Theorems 1.27–1.29.

14.1. Primitive induced maps and topological Markov shifts. Let $(\mathcal{Y}, \hat{F}_{\mathcal{Y}}, \tau)$ be a level- R primitive induced map of a piecewise-monotone map f with non-positive Schwarzian derivative. The connected components of the domain of $\hat{F}_{\mathcal{Y}}$ are countable and can be listed $\{\hat{X}_i\}_{i \in \mathcal{A}}$, for some countable alphabet \mathcal{A} . This gives a *natural coding* H defined on the invariant set

$$\Lambda := \bigcap_{n \geq 0} \hat{F}_{\mathcal{Y}}^{-n}(\mathcal{Y})$$

by $H(\hat{x}) = i_0 i_1 i_2 \dots$, where $\hat{F}_{\mathcal{Y}}^k(\hat{x}) \in \hat{X}_{i_k}$ for all $k \geq 0$. From Lemma 10.7 (and Lemma 4.15(v)), it follows that Λ is totally disconnected and the map H is injective. Note that, because \mathcal{Y} is a primitive, $\hat{F}_{\mathcal{Y}}$ is transitive on Λ . Set $\mathcal{X} := H(\Lambda)$ and denote by σ the left-shift on \mathcal{X} .

$$H \circ \hat{F}_{\mathcal{Y}} \circ H^{-1} = \sigma. \quad (14.1)$$

There is an *incidence matrix* associated to the shift; in Mauldin and Urbański's terminology, this matrix is irreducible, because $(\hat{F}_{\mathcal{Y}}$ and therefore) σ is transitive, while verification that the matrix is *finitely irreducible* is easy thanks to Lemma 4.15(i)-(iii). Summarising:

Lemma 14.1. *On the invariant set Λ , the coding map conjugates $\hat{F}_{\mathcal{Y}}$ to a Markov shift over a countable alphabet with a finitely irreducible incidence matrix (a finitely irreducible Markov shift).*

14.2. Induced Markov maps. The preceding construction can be made more general. Since we will need to deal with a second induced map, it is worthwhile to do so. Let (I, f) be some dynamical system defined in some metric space I equipped with a Borel σ -algebra, where f is a map with domain and range contained in I . We will call (Y, F, τ, π) an *expanding induced Markov map* for (I, f) if

- Y is a metric space with a finite number of connected components Y_1, \dots, Y_p ;
- on each connected component of Y , π is a metric isomorphism onto a subset of I ;
- the domain of definition of F can be written as $\bigcup_{i \in \mathcal{A}} X_i$ for some countable alphabet \mathcal{A} , where X_i are pairwise-disjoint subsets of Y , each compactly contained in Y and mapped homeomorphically by F onto some Y_j ;
- τ is constant on each X_i and $F(x) = f^{\tau(x)}(\pi(x))$;
- the maximal diameter of a connected component of $F^{-n}(Y)$ tends to 0 as $n \rightarrow \infty$.

We say (Y, F, τ, π) is *full-branched* if Y only has one connected component; in this case one can identify Y with $\pi(Y)$. Set $\Lambda(F) := \bigcup_{n \geq 0} F^{-n}(Y)$; this set is compact and totally disconnected. We say that (Y, F, τ) is *irreducible* if F is transitive on $\Lambda(F)$. If (Y, F, τ) is irreducible, then the natural coding map H is a conjugacy between F on $\Lambda(F)$ and a finitely irreducible Markov shift (\mathcal{X}, σ) over the alphabet \mathcal{A} .

14.3. Finitely irreducible Markov shifts. Let (\mathcal{X}, σ) be a finitely irreducible Markov shift over a countable alphabet \mathcal{A} . Denote by \mathcal{P}^σ the set of 1-cylinders (a 1-cylinder is a set of points of \mathcal{X} with the same first symbol) and by \mathcal{P}_n^σ the set of n -cylinders. Equip \mathcal{X} with the standard distance d , so the diameter of an n -cylinder is e^{-n} . Given (\mathcal{X}, σ) and potential $\Phi : \mathcal{X} \rightarrow \mathbb{R}$, we define the *n -th variation* by

$$V_n(\Phi) = V_n^\sigma(\Phi) := \sup_{\mathbf{Z} \in \mathcal{P}_n^\sigma} \sup_{x, y \in \mathbf{Z}} |\Phi(x) - \Phi(y)|, \tag{14.2}$$

We say that Φ is (C, α) -*locally Hölder continuous* (or just *Hölder continuous*) of order α if $V_n(\Phi) \leq C e^{-\alpha n}$ for some $C, \alpha > 0$, for all $n \geq 1$. ‘Locally’ indicates that one only compares values of the potential within 1-cylinders, which allows for unbounded potentials. We assume throughout this subsection that Φ is locally Hölder continuous.

For $A \subset \mathcal{A}$, let

$$\mathcal{X}_A := \{x \in \mathcal{X} : x_k \in A \text{ for all } k \geq 0\},$$

which is just the subset of \mathcal{X} consisting of symbolic sequences all of whose symbols belong to A . The notion of pressure has been extended to Markov shifts over countable alphabets, and in the current setting of finitely irreducible Markov shifts with a locally Hölder continuous potential,

$$P_\sigma(\Phi) = \sup \left\{ h(\mu) + \int \Phi d\mu \right\} \tag{14.3}$$

where the supremum can be taken over all measures $\mu \in \mathcal{M}_\sigma$ or over finite alphabets A and measures $\mu \in \mathcal{M}_\sigma$ supported on \mathcal{X}_A , see [MU, Theorem 2.1.5] and the top of page 10 in the same (recall \mathcal{M}_σ denotes the collection of ergodic σ -invariant probability measures). An equilibrium measure is an ergodic σ -invariant measure ν with $\int \Phi d\nu > -\infty$ for which $P_\sigma(\Phi) = h(\nu) + \int \Phi d\nu$.

If

$$\sum_{\mathbf{Z} \in \mathcal{P}^\sigma} \sup_{x \in \mathbf{Z}} \exp(\Phi(x)) < \infty,$$

then we say that Φ is *summable*. Note that a summable potential is necessarily bounded above (and unbounded below, if the alphabet is infinite).²

Remark 14.2. *By [MU, Proposition 2.1.9], the potential is summable if and only if the pressure is finite (summability is the same as $Z_1 < \infty$, by definition).*

The Birkhoff sum of Φ is

$$S_n^\sigma \Phi(x) := \sum_{j=0}^{n-1} \Phi \circ \sigma^j(x).$$

A measure ν is said to be a *Gibbs measure* for Φ (with *Gibbs constant* p) if there exists $K < \infty$ such that for all $\mathbf{Z} \in \mathcal{P}_n^F$,

$$\frac{1}{K} \leq \frac{\nu(\mathbf{Z})}{e^{S_n^\sigma \Phi(x) - np}} \leq K$$

for any $x \in \mathbf{Z}$. Given an expanding induced Markov map F and the natural coding map, if the push-forward of an F -invariant measure $\hat{\nu}$ under the coding map is a Gibbs measure, we also call $\hat{\nu}$ a Gibbs measure.

Lemma 14.3. *If Φ is locally Hölder continuous and is summable, then there exists a unique σ -invariant Gibbs measure ν_Φ for Φ . It has constant $p = P_\sigma(\Phi) := P(\sigma, \Phi)$ equal to the pressure. If $\int \Phi d\nu_\varphi > -\infty$ then ν_Φ is the unique equilibrium state for the potential Φ .*

Proof. The first statements follow from [MU, Corollary 2.7.5, Proposition 2.2.2], while for the final one we invoke [MU, Theorem 2.2.9]. \square

Denote by \mathcal{K}_β the set of functions $\Psi : \mathcal{X} \rightarrow \mathbb{C}$ which are locally Hölder continuous (defined in the same way as for real-valued functions) of order $\beta > 0$. Denote by \mathcal{K}_β^s the set of functions in \mathcal{K}_β whose real parts are summable, and by $L(\mathcal{K}_\beta)$ the space of bounded continuous operators on \mathcal{K}_β .

We define the transfer operator for Φ as

$$(\mathcal{L}_\Phi g)(x) := \sum_{\sigma(y)=x} e^{\Phi(y)} g(y),$$

for g in \mathcal{K}_β , so $\mathcal{L}_\Phi \in L(\mathcal{K}_\beta)$.

Lemma 14.4. *Let U be an open ball in \mathbb{C}^2 and let $\Psi_j \in \mathcal{K}_\beta$, $j = 0, 1, 2$. Suppose that for each $(u, v) \in U$, the potential $\Phi_{u,v} := \Psi_0 + u\Psi_1 + v\Psi_2 \in \mathcal{K}_\beta^s$, that is, each potential is summable. Then $(u, v) \mapsto \mathcal{L}_{\Phi_{u,v}}$ is analytic.*

Proof. From [MU, Corollary 2.6.10], $(u, v) \mapsto \mathcal{L}_{\Phi_{u,v}}$ is separately analytic. Then a vector-valued version of Hartogs' theorem [Col, 3.3.4] implies analyticity on U . \square

The following proposition is a two-variable version of [MU, Theorem 2.6.12], whose proof we follow.

²In [MU, Chapter 2] the recurring assumption that the potential Φ (there denoted f) is bounded ought to read that $\exp(\Phi)$ is bounded.

Proposition 14.5. *Let U be an open ball in \mathcal{C}^2 and let $\Psi_j \in \mathcal{K}_\beta$, $j = 0, 1, 2$. Suppose that for each $(u, v) \in U$, the potential $\Psi_{u,v} := \Psi_0 + u\Psi_1 + v\Psi_2 \in \mathcal{K}_\beta^s$. If A is (in the terminology of [MU]) finitely primitive (in particular, if σ is a full shift, or finitely irreducible and topologically mixing), then*

$$(u, v) \mapsto P_\sigma(\Psi_{u,v})$$

is analytic.

Proof. By [MU, Theorem 2.4.6] (which has ‘finitely primitive’ as a hypothesis), $\mathcal{L}_{\Psi_{u,v}}$ has a simple isolated eigenvalue equal to $\exp(P(\Psi_{u,v}))$, and analytic perturbation theory then implies that this eigenvalue moves analytically with (u, v) , from which the result follows. \square

14.4. Pressure above and below, I. As in §14.2, let (Y, F, τ) be an induced Markov map for some dynamical system (I, f) . Let the coding map H denote the conjugacy with (\mathcal{X}, σ) , the corresponding finitely irreducible Markov shift. Suppose $\hat{\nu} \in \mathcal{M}_F$ spreads to $\mu \in \mathcal{M}_f$ and let ν denote $H_*\hat{\nu} \in \mathcal{M}_\sigma$. Given $\varphi : I \rightarrow \mathbb{R}$ an integrable potential function, its corresponding induced potential $\Phi : \mathcal{X} \rightarrow \mathbb{R}$ is

$$\Phi := S_\tau \varphi \circ \pi \circ H^{-1} := \sum_{j=0}^{n-1} \varphi \circ f^j \circ \pi \circ H^{-1}.$$

Write $T := \int_Y \tau d\hat{\nu}$. The entropies satisfy Abramov’s formula (8.4), so $h(\mu) = T^{-1}h(\nu)$. By ergodicity, Φ satisfies (similarly to (8.5))

$$\int_I \varphi d\mu = \frac{1}{T} \int_{\mathcal{X}} \Phi d\nu. \quad (14.4)$$

Lemma 14.6. *The free energies satisfy*

$$E(f, \varphi, \mu) = \frac{1}{T} E(\sigma, \Phi, \nu).$$

In particular, $E(f, \varphi, \mu)$ has the same sign (positive, negative or zero) as $E(\sigma, \Phi, \nu)$.

Proof. This follows immediately from Abramov’s formula and (14.4). \square

Lemma 14.7 ([IT1, Lemma 4.1]). *Suppose $P(f, \varphi) = 0$ and Φ is locally Hölder continuous. Then $P_\sigma(\Phi) \leq 0$.*

Proof. By (14.3), if $P_\sigma(\Phi) > 0$, then there is a measure $\nu \in \mathcal{M}_\sigma$ with $E(\sigma, \Phi, \nu) > 0$ and supported on \mathcal{X}_A for some finite alphabet A . On $H^{-1}(\mathcal{X}_A)$, τ is bounded, so ν spreads to a measure $\mu \in \mathcal{M}_f$. By Lemma 14.6, $E(\mu, \varphi)$ is strictly positive, contradicting the definition of pressure as the supremum of free energies. \square

14.5. Pressure above and below, II. Let us return to our specific maps and potentials. Let $(\mathcal{Y}, \hat{F}_\mathcal{Y}, \tau)$ be a level- R primitive induced map of a d -branched piecewise-monotone map f with non-positive Schwarzian derivative. Let (\mathcal{X}, σ) be the corresponding finitely irreducible Markov shift and H the conjugacy between the two maps from (14.1). We consider the potential

$$\varphi_{t,v} := -t \log |Df| - v.$$

The corresponding induced potential on \mathcal{X}

$$\Phi_{t,v} := S_\tau \varphi_{t,v} \circ \pi \circ H^{-1} \quad (14.5)$$

satisfies

$$\Phi_{t,v} \circ H = -t \log |D\hat{F}_Y| - v\tau.$$

From Lemma 10.10, $\Phi_{t,v}$ is locally Hölder continuous (note τ is constant on cylinders, so $V_n(\tau) = 0$).

Suppose $\hat{\nu} \in \mathcal{M}_{\hat{F}_Y}$ spreads to $\mu \in \mathcal{M}_f$ and let ν denote $H_*\hat{\nu} \in \mathcal{M}_\sigma$. From Lemma 14.6, $E(f, \varphi_{t,v}, \mu)$ has the same sign (positive, negative or zero) as $E(\sigma, \Phi_{t,v}, \nu)$.

Let us set $p(t) := P(-t \log |Df|)$. Immediately, $P(\varphi_{t,p(t)}) = 0$. From Lemma 14.7, we obtain:

Lemma 14.8. $P_\sigma(\Phi_{t,p(t)}) \leq 0$.

Lemma 14.9. *Suppose that $\mu \in \mathcal{M}_f$ is an equilibrium measure for a potential $-t \log |Df|$ with $\hat{\mu}(\mathcal{Y}) > 0$. Then $P_\sigma(\Phi_{t,p(t)}) = 0$. The potential $\Phi_{t,p(t)}$ is summable. The corresponding induced measure in \mathcal{M}_σ*

$$\nu = H_* \frac{1}{\hat{\mu}(\mathcal{Y})} \hat{\mu}|_{\mathcal{Y}} \tag{14.6}$$

is the unique equilibrium measure for $\Phi_{t,p(t)}$ and is a Gibbs measure.

Proof. We have $E(f, \varphi_{t,p(t)}, \mu) = 0$, hence (by Lemma 14.6) the corresponding measure in \mathcal{M}_σ has zero free energy for the potential $\Phi_{t,p(t)}$. Therefore $P_\sigma(\Phi_{t,p(t)}) \geq 0$ which, combined with Lemma 14.8, gives the first statement. Since the pressure is finite (equal to 0), the potential is summable by Remark 14.2. The remaining statement now follows from Lemma 14.3 and (14.3). \square

Proof of Theorem 1.27. Let $d \geq 2, \varepsilon > 0$. Take R sufficiently large that $2\varepsilon(R, d) < \varepsilon$. By Remark 4.13, there are at most $(2dR)^2 d^R$ distinct level- R primitive induced maps. If f is transitive on $J(f)$, Lemma 4.14 says there is only one level- R primitive induced map. There is at most one equilibrium state which lifts to a given primitive map, by Lemma 14.9, and each equilibrium state with entropy at least $2\varepsilon(R)$ does lift to such a map, by Corollary 8.3. Thus the number of level- R primitive induced maps bounds the number of equilibrium states with entropy at least ε , completing the proof. \square

14.6. Exponential tails. We show existence of a full-branched map with exponential tails. Let f be a d -branched piecewise-monotone map with non-positive Schwarzian derivative and set $p(t) := P(-t \log |Df|)$.

Theorem 14.10. *Given $\theta > 0$, there is a finite collection \mathcal{K} of full-branched expanding induced Markov maps for f satisfying the following. Suppose*

$$p(t) - \max\{-t\lambda_{\min}, -t\lambda_{\max}\} > 2\theta,$$

so f has a positive-entropy equilibrium state μ for the potential $-t \log |Df|$. For some $(Y_0, G, \tau_0, \pi) \in \mathcal{K}$, the measure μ lifts to a measure $\hat{\nu}_0 \in \mathcal{M}_G$, and

- $\hat{\nu}$ is a Gibbs measure for the potential $S_{\tau_0} \varphi_{t,p(t)} \circ \pi$ with Gibbs constant 0;
- \log of the Jacobian of $\hat{\nu}_0$ is locally Hölder;
- the map is κ -extendible for some $\kappa > 0$: if $\tau_0 = n$ on a 1-cylinder V of G , then f^n maps a neighbourhood of $\pi(V)$ onto the κ -neighbourhood of $\pi(Y_0)$;
- $|DG| > 2$;

- there exists constants $C_2, \theta_0 > 0$ for which

$$\hat{\nu}_0(\tau_0^{-1}(n)) \leq C_2 \exp(-n\theta_0)$$

for all $n \geq 1$.

Moreover, if f is transitive on $J(f)$, one can take $\#\mathcal{K} = 1$.

Proof. Given $\theta > 0$, take R large enough that $\varepsilon(R) < \theta$. There are only finitely many level- R primitive induced maps, and only one when f is transitive. For each such induced map $(\mathcal{Y}, \hat{F}_{\mathcal{Y}}, \tau)$, for K large enough $|D\hat{F}_{\mathcal{Y}}^n| > 2$ for all $n \geq K$. Select some K -cylinder Y_0 of $\hat{F}_{\mathcal{Y}}$. The first return map G to Y_0 is full-branched, expanding and extendible. Note that G is the first return map for \hat{f} and for $\hat{F}_{\mathcal{Y}}$; we choose the return time τ_0 to be with respect to \hat{f} . We thus obtain a finite collection \mathcal{K} of maps (Y_0, G, τ_0, π) , each one corresponding to a primitive induced map. We shall obtain tail estimates for $(\mathcal{Y}, \hat{F}_{\mathcal{Y}}, \tau)$ and subsequently transfer them to the corresponding (Y_0, G, τ_0, π) .

Suppose $t \in \mathbb{R}$ satisfies

$$p(t) - \max\{-t\lambda_{\min}, -t\lambda_{\max}\} > 2\theta. \quad (14.7)$$

By Corollary 1.20, there exists an equilibrium measure μ for the potential $-t \log |Df|$. The entropy of μ is necessarily at least 2θ , by (14.7). Since $h(\mu) > 2\varepsilon(R)$, $\mu(\mathcal{Y}) > 0$ for some level- R primitive induced map $(\mathcal{Y}, \hat{F}_{\mathcal{Y}}, \tau)$. Obtain the conjugacy H to (\mathcal{X}, σ) as before. Let $\varphi_{t,v}, \Phi_{t,v}$ be defined as per (14.5). Let

$$\hat{\nu} := \frac{1}{\hat{\mu}(\mathcal{Y})} \hat{\mu}|_{\mathcal{Y}}$$

and set $\nu := H_*\hat{\nu}$, so ν is the corresponding equilibrium measure ν for σ given by (14.6). By Lemma 14.9, ν is a Gibbs measure for $\Phi_{t,p(t)}$. By Lemma 10.4,

$$-t \log |D\hat{F}_{\mathcal{Y}}| - \tau p(t) \leq -2\theta\tau + C,$$

for some constant $C > 0$. By the Gibbs property of ν , for each 1-cylinder $\mathbf{Z} \subset \mathcal{Y}$ of $\hat{F}_{\mathcal{Y}}$ with $\tau(\mathbf{Z}) = n$,

$$\begin{aligned} \hat{\nu}(\mathbf{Z}) &\leq K \sup_{\mathbf{Z}} \exp\left(-t \log |D\hat{F}_{\mathcal{Y}}| - p(t)\tau\right) \\ &\leq K \exp(-2\theta\tau + C). \end{aligned}$$

In Lemma 4.15, we gave a counting estimate on the number of cylinders with a given inducing time,

$$\#\{\mathbf{Z} : \tau(\mathbf{Z}) = n\} \leq C_0 \exp(\varepsilon(R)n),$$

for some $C_0 > 0$ and all $n \geq 1$. Combining the two estimates, we obtain

$$\hat{\nu}(\tau^{-1}(n)) \leq KC_0 e^C \exp(-n\theta)$$

a basic exponential tails estimate from which much follows. We use this to obtain exponential tails for a full-branched induced map.

Let $V = X_{i_1 i_2 \dots i_p}$ denote the set of points for which

$$\hat{F}_{\mathcal{Y}}^j(V) \subset \tau^{-1}(i_j)$$

for $j = 1, 2, \dots, p$. If $n := \sum_{j=1}^p i_j$, then

$$\hat{\nu}(V) \leq K \exp(-2\theta n + C) C_0^p \exp(\varepsilon(R)n)$$

so

$$\nu(V) \leq C_1^p \exp(-n\theta), \quad (14.8)$$

for some $C_1 > 1$ and all n .

Let (Y_0, G, τ_0, π) be the full-branched induced map corresponding to \mathcal{Y} . By the Gibbs property, Y_0 has positive measure. Let $\hat{\nu}_0 \in \mathcal{M}_G$ denote the normalised restriction of $\hat{\nu}$ to Y_0 . The Gibbs property of $\hat{\nu}_0$ is inherited from $\hat{\nu}$; alternatively, one could reproduce §14.5. That log of the Jacobian is locally Hölder follows from the Gibbs property and bounded distortion.

We first consider points which make many returns to \mathcal{Y} before returning to Y_0 . By Lemma 4.15(iii), there exists N such that for every 1-cylinder \mathbf{Z} of $\hat{F}_{\mathcal{Y}}$,

$$\mathcal{Y} \subset \bigcup_{j=1}^N \hat{F}_{\mathcal{Y}}^j(\mathbf{Z}).$$

Together with the Gibbs property and bounded distortion, this implies the existence of some $\beta > 0$ such that, given an n -cylinder \mathbf{Z} ,

$$\hat{\nu} \left(\{ \hat{x} \in \mathbf{Z} : \hat{F}_{\mathcal{Y}}^j(\hat{x}) \notin Y_0, \quad j = n, \dots, n+N-1 \} \right) \leq \exp(-\beta).$$

Since the complement of $\{ \hat{x} \in \mathbf{Z} : \hat{F}_{\mathcal{Y}}^j(\hat{x}) \notin Y_0, \quad j = n+1, \dots, n+N \}$ is a union of $(n+N+K)$ -cylinders, we can inductively obtain

$$\hat{\nu} \left(\{ \hat{x} \in Y_0 : \hat{F}_{\mathcal{Y}}^j(\hat{x}) \notin Y_0, \quad j = 1, \dots, n(N+K) \} \right) \leq \exp(-n\beta). \quad (14.9)$$

For $T \leq n$, we can split the set $\tau_0^{-1}(n)$ in two depending on whether $\tau_0 \geq S_T^\sigma \tau$ or not (that is, whether a point making its first return to Y_0 has passed more or fewer than T times through \mathcal{Y}). From (14.9) and (14.8) we deduce that

$$\begin{aligned} \hat{\nu}(\tau_0^{-1}(n)) &\leq \sum_{p=T}^n \exp(-\lfloor p/N \rfloor \beta) + \sum_{p=1}^T \sum_{i_1+\dots+i_p=n} C_1^p \exp(-n\theta) \\ &\leq C_3 \exp(-T\beta/N) + C_1^T 2^T \exp(-n\theta), \end{aligned}$$

for some $C_3 > 1$. If one takes $T := \lfloor \gamma n \rfloor$ with $\gamma := \theta/2 \log C_1 2$, each term is exponentially small in n . Therefore $\hat{\nu}_0(\tau_0^{-1}(n))$ is also exponentially small in n , as required. \square

14.7. Analyticity of pressure.

Proof of Theorem 1.28. We assume $t \in (t^-, t^+)$, so

$$p(u) - \max\{-u\lambda_{\min}, -u\lambda_{\max}\} > 2\theta$$

for some $\theta > 0$, for all u in a small neighbourhood of t . Moreover f is assumed transitive, so we have a unique system (Y_0, G, τ_0, π) given by Theorem 14.10. To each equilibrium state μ_u for $-u \log |Df|$, we obtain $\hat{\nu}_0^u \in \mathcal{M}_G$ also from the theorem.

The system has an invariant set

$$\Lambda_0 := \bigcap_{n \geq 0} G^{-n}(Y_0).$$

We obtain a new coding map $H_0 : \Lambda_0 \rightarrow \mathcal{X}_0$ and conjugacy $H_0 \circ G \circ H_0^{-1} = \sigma_0$, where $(\mathcal{X}_0, \sigma_0)$ is the full shift. Let $\nu_0^u := (H_0)_* \hat{\nu}_0^u$. Analogously to (14.5), we set $\varphi_{u,v} := -u \log |Df| - v$ and define the potential

$$\Psi_{u,v} := S_{\tau_0} \varphi_{u,v} \circ \pi \circ H_0^{-1}, \quad (14.10)$$

Now ν_0^u is a σ_0 -invariant Gibbs measure for the potential $\Psi_{u,p(u)}$ with Gibbs constant 0. From Theorem 14.10, ν_0^t has exponential tails: There exist constants $C_2, \theta_0 > 0$ for which

$$\nu_0^t(\tau_0^{-1}(n)) \leq C_2 \exp(-n\theta_0) \tag{14.11}$$

for all $n \geq 1$. From the Gibbs property of the measure, we recover immediately from (14.11) estimates on the potential: For some $\theta_0 > 0$,

$$\sum_{\mathbf{Z} \in \mathcal{P}^{\sigma_0}, \tau_0(\mathbf{Z})=n} \sup_{x \in \mathbf{Z}} \exp(\Psi_{t,p(t)}(x)) = O(e^{-\theta_0 n}).$$

From this, it immediately follows that $\Psi_{t,p(t)}$ is summable and moreover that, for all $(u, v) \in \mathcal{C}^2$ in a complex ball U containing $(t, p(t))$, $\Psi_{u,v}$ is in \mathcal{K}_β^s . Since σ_0 is a full shift, applying Proposition 14.5 produces: $(u, v) \mapsto P_{\sigma_0}(\Psi_{u,v})$ is analytic on U .

Let $U^\mathbb{R} := \Re(U)$ be a real ball containing $(t, p(t))$ on which $(u, v) \mapsto P_{\sigma_0}(\Psi_{u,v})$ is real-analytic. For $(u, v) \in U^\mathbb{R}$, $\Psi_{u,v}$ is real-valued and summable, so Lemma 14.3 implies the existence of a Gibbs measure $\nu_{u,v}$ for $\Psi_{u,v}$. By the variational definition of pressure (14.3), consideration of the free energy of $\nu_{u,v}$ gives, in $U^\mathbb{R}$,

$$\left. \frac{\partial P_{\sigma_0}(\Psi_{u,v})}{\partial v} \right|_{u,v} = - \int \tau \, d\nu.$$

In particular for each u , there is at most one v with $P_{\sigma_0}(\Psi_{u,v}) = 0$.

Consideration of ν_0^u and Lemma 14.3 gives $P_{\sigma_0}(\Psi_{u,p(u)}) = 0$ on a neighbourhood of t . By the Implicit Function Theorem, we deduce that $u \mapsto p(u)$ is analytic at $u = t$. This holds for each $t \in (t^-, t^+)$, completing the proof of Theorem 1.28. \square

14.8. ASIP and decay of correlations.

Proof of Theorem 1.29. Here we will apply the results of [Y, MN] to prove Theorem 1.29. Let $t \in (t^-, t^+)$; there is a corresponding equilibrium state μ for the potential $-t \log |Df|$. Obtain (Y_0, G, τ_0, π) from Theorem 14.10. It is a full-branched expanding induced Markov map with a G -invariant Gibbs measure $\hat{\nu}_0$ with exponential tails which spreads to μ . In particular, $\hat{\nu}_0(\tau \geq n) = O(e^{-\eta n})$. From the κ -extendibility of G and non-positive Schwarzian derivative, there exists $C > 0$ so that for x, y in the same connected component V of the domain of G , and identifying x, y with $\pi(x), \pi(y)$,

$$|f^k(x) - f^k(y)| \leq C |G^R(x) - G^R(y)| \text{ for } 1 \leq k \leq R - 1 \tag{14.12}$$

where $R = \tau(x) = \tau(y)$.

We apply [MN, Theorem 2.9], noting that conditions (1)–(5) in [MN, §2(e)] hold for our system: (1) (G is full-branched); (2) (G is uniformly expanding); (3) (inequality (14.12)); (4) (the log of the Jacobian of $\hat{\nu}_0$ is locally Hölder); (5) (integrable inducing time). We obtain that mean-zero Hölder observations satisfy the ASIP, as required.

The induced Markov map can be viewed as a Young tower [Y]. One applies [Y, Theorem 3] to establish decay of correlations ([Y, §6.4] shows how to go back from the tower to the original system). \square

15. INSTABILITY FOR COLLET-ECKMANN MAPS

The class of unimodal maps with exponential growth along critical orbits is usually considered to have almost as good statistical properties as uniformly hyperbolic maps. However, in this section we will prove Theorem 1.30, showing that these properties do not extend to statistical stability. We no longer deal with potentials, freeing up the use of the notation φ, ψ for some induced maps.

Let \mathcal{F} denote the set of C^2 unimodal maps $f : I \rightarrow I$ with non-positive Schwarzian derivative, $f(\partial I) \subset \partial I$ and $|Df|_{\partial I}| > 1$, where I denotes the compact interval $[0, 1]$. Let us endow \mathcal{F} with the C^0 topology. Each map f has a unique turning (critical) point c in the interior of I .

In §15.2 we will show that, for a map in \mathcal{F} with a parabolic orbit and a certain sequence of Misiurewicz maps converging to the parabolic map, the acips for the Misiurewicz maps converge to the equidistribution along the parabolic orbit. We require stronger control over the acips than mere existence, so we provide another proof of the existence of such measures, while keeping some control on the densities; nevertheless, our proof is a little simpler than the comparable proof in [BM].

In §15.3, we examine conjugacies from unimodal maps to tent maps and show existence of tent maps with certain combinatorial properties.

In the final subsection, we conclude the proof of Theorem 1.30. In particular, we show that once the topological entropy varies in a continuous family in \mathcal{F} , starting from a non-renormalisable map, there are parabolic maps converging to that map for which the equidistributions along the parabolic period orbits converge to a given periodic orbit of the initial map. Moreover, we show that sequences of Misiurewicz maps as considered in Subsection 15.2 exist. Taking a diagonal subsequence gives a sequence of Misiurewicz maps whose acips converge to the given periodic orbit.

Given an interval J and a differentiable map $h : J \rightarrow \mathbb{R}$, we say that h has distortion bounded by C if

$$\sup_{x, y \in J} |Dh(x)|/|Dh(y)| \leq C. \quad (15.1)$$

If h has bounded distortion clearly it is a diffeomorphism. If h is a diffeomorphism with *non-positive Schwarzian derivative* on a larger interval $J' \supset J$, then the Koebe Principle gives an *a priori* distortion bound for h on J which depends solely on the relative lengths of $h(J)$ and the connected components of $h(J') \setminus h(J)$. If the domain of h is instead a collection of pairwise-disjoint intervals, we will say that h has distortion bounded by C if (15.1) holds on each connected component of its domain.

We denote by (a, b) the open interval with boundary points a, b , regardless of whether $a < b$ or vice versa, and by $[a, b]$ the corresponding closed interval.

15.1. Preliminary lemmas. The following lemmas will be useful to quantify properties of the acips we construct in the proof of Proposition 15.6.

Lemma 15.1. *Let a_k satisfy $0 < a_k < 1$ and $\sum_{k=1}^{\infty} a_k k \leq C < \infty$. Then $\sum_{k=1}^{\infty} -a_k \log a_k < C + 12$.*

Proof. Split the sum into two parts, one where $-\log a_k \leq k$, where the sum is bounded by C , and a second part where $a_k < \exp(-k)$. Now $-\log r \leq 2\sqrt{1/r}$ for $r > 0$, so $-a_k \log a_k \leq 2\sqrt{a_k} < 2\exp(-k/2)$ provided $a_k < \exp(-k)$. The expression

$$2 \sum_{k \geq 1} \exp(-k/2) < 2(1 + \exp(1/2)) \sum_{k \geq 1} \exp(-k)$$

is bounded by 12. \square

Lemma 15.2. *Let $N, \Delta, \lambda > 1, \delta > 0$ be given. Let I be an interval of length 1. Then there exists $\beta > 0$ such that the following holds. Let U be an open subinterval of I of length $\geq \delta$. Let $f : I \rightarrow I$ be a C^2 map such that*

•

$$\text{dist} \left(U, \bigcup_{i \geq 1} f^i(\partial U) \right) \geq \delta;$$

- $|Df^n(x)| \geq \lambda$ for all $n \geq N$ and x satisfying $f^j(x) \notin U$ for $j = 0, 1, \dots, n-1$;
- $\inf\{|Df(x)| : x \notin U\} \geq \delta$ and $f(\partial I) \subset \partial I$;
- given any interval V and n such that $V, f(V), \dots, f^{n-1}(V)$ are each disjoint from U : then the distortion of f^j on V for $j = 1, \dots, n$ is bounded by Δ .

Let $e(x)$ denote the first entry time of x into U , that is, $e(x) = \inf\{k \geq 0 : f^k(x) \in U\}$ (if $f^k(x) \notin U$ for all $k \geq 0$, then $e(x) = \infty$). For all $n \geq 0$,

$$m(\{x : e(x) \geq n\}) \leq \exp(-n\beta).$$

Proof. Mimic the proof of [MeS, Lemma V.3.3]. \square

Lemma 15.3. *For each $\beta > 0$ there is a constant C such that the following holds. Let I be an interval and r a measurable function from I to \mathbb{N} such that, for each $k \geq 0$, $m(\{x \in I : r(x) = k\}) \leq \exp(-k\beta)$.*

Let $J \subset I$ be a measurable set of measure bounded by $\gamma > 0$. Then $\int_J r(x) dx \leq C\gamma(C - \log \gamma)$

Proof. For integers n , let $S_n := \sum_{k=n}^{\infty} \exp(-k\beta) = \exp(-n\beta)/(1 - \exp(-\beta))$. By examining the worst case scenario, it is easy to see that the integral of r over any set of measure bounded by S_n is bounded by $T_n := \sum_{k=n}^{\infty} k \exp(-k\beta)$.

Let N be maximal such that $\gamma \leq S_N$. Then rearranging $S_{N+1} \leq \gamma \leq S_N$ gives

$$\frac{\log(\gamma(1 - \exp(-\beta)))}{-\beta} - 1 \leq N \leq \frac{\log(\gamma(1 - \exp(-\beta)))}{-\beta} + 1.$$

Now, $\int_J r(x) dx \leq T_N$. We have

$$\begin{aligned} T_N &= \left(N + \frac{e^{-\beta}}{1 - e^{-\beta}} \right) \frac{e^{-N\beta}}{1 - e^{-\beta}} \\ &\leq \left(\frac{\log(\gamma(1 - e^{-\beta}))}{-\beta} + 1 + \frac{e^{-\beta}}{1 - e^{-\beta}} \right) \frac{e^\beta \gamma (1 - e^{-\beta})}{1 - e^{-\beta}}, \end{aligned}$$

which can be written in the desired form for some constant C . \square

For motivation, one can think of h in the following lemma as a unimodal map f restricted to a one-sided neighbourhood of the critical point, and of r as the first entry time (on a neighbourhood of the critical value) to a neighbourhood of the critical point. Then this lemma will control integral of the return time on a critical neighbourhood.

Lemma 15.4. *For each pair of constants $\beta, C_1 > 0$ there is a constant L such that the following holds. Let h be a C^2 diffeomorphism from an open interval U onto an interval $W := h(U)$, where both intervals have length bounded by 1. Suppose $|Dh| < C_1$ and $-\int_U \log |Dh(x)| dx < C_1$. Let r be a measurable function from W to \mathbb{N} such that, for each $k \geq 0$, $m(\{x \in I : r(x) = k\}) \leq \exp(-k\beta)$.*

Then $\int_U r(h(x)) dx < L$.

Proof. Let C be given by Lemma 15.3, so if J is a set with measure bounded by $\gamma > 0$, then

$$\int_J r(x) dx \leq C\gamma(C - \log \gamma). \quad (15.2)$$

For each $k \geq 0$, let U_k denote the set of $x \in U$ for which $e^{-k} \geq |Dh(x)|/C_1 > e^{-k+1}$. Then $U = \bigsqcup_{k=0}^{\infty} U_k$, and

$$C_1 > -\int_U \log |Dh| dx \geq -\log C_1 + \sum_{k=0}^{\infty} km(U_k).$$

Now, $\int_{U_k} r(h(x)) dx \leq e^{k+1} \int_{h(U_k)} r(y) dy$. We have $m(h(U_k)) \leq e^{-k} m(U_k)$, so (15.2) gives

$$\int_{h(U_k)} r(y) dy \leq Ce^{-k} m(U_k)(C - \log(e^{-k} m(U_k))) \leq Ce^{-k} m(U_k)(C + k - \log m(U_k)).$$

Then

$$\int_{U_k} r(h(x)) dx \leq C e m(U_k)(C + k - \log m(U_k)).$$

Summing over k and using Lemma 15.1 with $a_k = m(U_k)$, we get,

$$\int_U r(h(x)) dx \leq Ce(m(U)C + (C_1 + \log C_1) + C_1 + 12).$$

We can choose $L := Ce(C + 2C_1 + \log C_1 + 12)$. □

15.2. Acips for some almost-parabolic Misiurewicz maps. Let $g \in \mathcal{F}$ have a parabolic periodic orbit of (minimal) period k , and suppose that g^k is orientation-preserving in a neighbourhood of the orbit. Let α be the point of the orbit such that the critical point c (of g) is in the immediate basin of attraction of α for the iterate g^k ([MeS, Theorem III.6.1]). Let c' be the nearest critical point of g^k on the other side of α from c , so for some $j < k$, $g^j(c') = c$. Then g^k is orientation-preserving on (c', c) .

Lemma 15.5. *The graph of g^k does not traverse the diagonal on (c', c) and α is the unique fixed point of g^k in this interval.*

Proof. If there were a second fixed point α' of g^k in (c', c) , one could use the Minimum Principle to show that it must be hyperbolic attracting. Thus it would not be in the orbit (under g) of α and c would be in the basin of two different attracting orbits, which would be impossible. It remains to rule out the case that c and c' are on opposite sides of the diagonal. Assume that they are on opposite sides, so α is attracting on both sides. Then for n large, $g^{nk}((c', c))$ is a

small neighbourhood of α . But for some $0 < j < k$, $g^j(c') = c$. Thus $g^{nk+j}(\alpha) = g^j(\alpha)$ is close to α and so equals α , contradicting minimality of k . \square

Let q be a preimage of some repelling periodic point so that g^k is monotone on (q, α) and $g^i(q), g^i(\alpha) \notin (q, \alpha)$ for all $i \geq 1$. It follows that $\alpha \in (q, c)$, since all points in (α, c) are in the basin of attraction of α .

Given a point $x \in I$ and $f \in \mathcal{F}$, we denote by x^* the other ('symmetric') point satisfying $f(x) = f(x^*)$. For maps close enough to g , there are corresponding pre-periodic points to q and q^* and we denote these also by q, q^* , without indicating the dependence on the map. The critical point is denoted by c , again without indicating dependence on the map.

Proposition 15.6. *Suppose there are $g_n \in \mathcal{F}$ converging to g in the C^0 topology such that $g_n^{ik}(c) \in (q, c)$ for $1 \leq i < n$ and $g_n^{nk}(c) = q$, and such that there exists $C > 0$ with $\int \log |Dg_n(x)| dx > -C$ for all large n .*

Then, for large n , each g_n has an acip μ_n and the sequence of measures μ_n converges weakly to the equidistribution on the orbit of α .

Proof. Assume n is large, so g_n is close to g .

Define q_i , for $1 \leq i < n$, by $q_i \in (q, c)$, $g_n^k(q_1) = q$ and $g_n^k(q_{i+1}) = q_i$. Then $g_n^k(c) = q_{n-1}$. Set $U := (q_{n-1}, q_{n-1}^*)$. Note that $\limsup_{n \rightarrow \infty} |c - q_{n-1}| > 0$.

The *first entry time* $e_V(x)$ of a point x to $V := (q_2, q_2^*)$ is defined as

$$e_V(x) := \inf\{j \geq 0 : g_n^j(x) \in V\}.$$

Define the *first entry map* φ_V to V by $\varphi_V(x) = g_n^{e_V(x)}(x)$. Now V is a nice interval containing the unique critical point, and q never returns to (q, q_1^*) , by choice of q . Then on each connected component of the domain of definition of φ_V , the restriction of φ_V is a diffeomorphism onto V which extends to a larger domain and this larger domain gets mapped by the corresponding iterate of g_n diffeomorphically onto (q, q_1^*) . Of course, φ_V is just the identity on $V = (q_2, q_2^*)$.

Set $W := (q_1, q_2)$ and define the corresponding first entry time e_W and first entry map φ_W to W . Note that $e_W(x) = (i-1)k$ on $(q_i, q_i^*) \setminus [q_{i+1}, q_{i+1}^*]$ for $2 \leq i < n-1$. By non-positive Schwarzian derivative, there is a uniform bound Δ , independent of large n , on the distortion of each of φ_V , $\varphi_W|_{(V \setminus U)}$ and $\psi := \varphi_V \circ \varphi_W$.

By the Mañé Hyperbolicity Theorem ([MeS, Theorem III.5.1]), g is uniformly hyperbolic away from V . Therefore the same holds for g_n for all large n : there exist $\lambda, j > 1$ such that $|D(g_n|_{(I \setminus V)})^j| > \lambda$.

Define $r(x)$ by $r := e_W + e_V \circ \varphi_W$. On V , $\psi(x) = g_n^{r(x)}(x)$. By the Folklore Theorem, there is an acip ν_n for $\psi : V \rightarrow V$ (ψ defined almost everywhere) and, because of the distortion bound,

$$\Delta^{-1} \leq \frac{d\nu_n}{dx} \cdot |V| \leq \Delta \tag{15.3}$$

on V . One can then define an absolutely continuous, g_n -invariant measure μ_n by, writing χ_S for the characteristic function of a set S ,

$$\mu_n(S) = \int_V \sum_{j=0}^{r(x)-1} \chi_S(g_n^j(x)) d\nu_n(x)$$

for all measurable sets S . Of course, μ_n is finite if and only if the integral is finite, with I in place of S — this is just $\int_V r(x) d\nu_n(x)$. We shall show finiteness and more now.

Our goal is to show that, for large n , most of μ_n is concentrated in a small neighbourhood of the orbit of α . So let B be an ε -neighbourhood of the orbit of α . It suffices to prove that $\mu_n(B)$ and $\mu_n(I \setminus B)$ are finite (showing finiteness of $\mu_n(I)$) and that $\lim_{n \rightarrow \infty} \mu_n(B)/\mu_n(I \setminus B) = \infty$.

For some M , for all large n , $\#\{i \leq e_W(x) : g_n^i(x) \notin B\} < M$ for all $x \in V$. Set $U_j := (q_j, q_{j+1}) \cup (q_j^*, q_{j+1}^*)$ for $j = 2, \dots, n-2$; on each U_j , $e_W(x) = kj$, and $V = U \cup \bigcup_{j=2}^{n-2} U_j$. Then,

$$\int_V e_W(x) d\nu_n(x) \geq \mu_n(B) = \int_V \sum_{j=0}^{r(x)-1} \chi_B(g_n^j(x)) d\nu_n(x) \geq \int_V (e_W(x) - M) d\nu_n(x).$$

Using (15.3) we deduce that

$$|V|\Delta k \left(n|U| + \sum_{j=2}^{n-2} |U_j|j \right) \geq \mu_n(B) \geq |V|\Delta^{-1} \left(-M|V| + k \sum_{j=2}^{n-2} |U_j|j \right).$$

Clearly $\mu_n(B)$ is finite. Denoting by U_j^n the set U_j for the map g_n , we have the following. For each j , the sets U_{n-j}^n converge as n tends to infinity to a set U_j^∞ , say, of definite size. Therefore, the term $\sum_{j=2}^{n-2} |U_j^n|j \geq |U_{n-2}^n|(n-2)$ tends to infinity with n , forcing $\mu_n(B)$ along with it.

It remains to show that $\mu_n(I \setminus B)$ is bounded independently of n . Now $\mu_n(I \setminus B) \leq \int_V (M + e_V(g^{e_W}(x))) d\nu_n(x)$. Again using (15.3),

$$\mu_n(I \setminus B) \leq |V|\Delta \left(|V|M + \Delta \sum_{j=2}^{n-2} |U_j| \int_W e_V(x) dx / |W| + \int_U e_V(g_n^{kn}(x)) dx \right).$$

By Lemma 15.2 there is a $\theta > 0$ such that $m(\{x \in I : e_V(x) = j\}) \leq \exp(-\theta j)$, again independently of n , provided n is large. Therefore $\int_W e_V(x) dx$ is bounded. Now consider the term $\int_U e_V(g_n^{kn}(x)) dx$. On U , g_n is a unimodal map; $g_n^{n^{k-1}}$ maps $g_n(U)$ diffeomorphically with distortion bound Δ and derivative bounded away from 0 and infinity independently of n , into W . On W , the return time e_V to V decays exponentially by Lemma 15.2, and the constants do not depend on n , so we can apply Lemma 15.4. Note we do not apply it to g_n^{nk} directly, but rather to $h := g_n^{nk}$ restricted to the two connected components of $U \setminus \{c\}$ one at a time. Applying the lemma gives a bound independent of n on $\int_U e_V(g_n^{kn}(x)) dx$. Therefore $\mu_n(I \setminus B)$ is uniformly bounded and $\mu_n(B)/\mu_n(I \setminus B)$ tends to infinity. Normalise the measure μ_n to finish. \square

15.3. Conjugacies to tent maps.

Definition 15.7. For $s \in (1, 2]$ denote by $T_s : I \rightarrow I$ the tent map with slope $\pm s$, defined by

- for $0 \leq x \leq 1/2$, $T_s(x) = sx$;
- for $1/2 \leq x \leq 1$, $T_s(x) = s - sx$.

The tent map T_s has topological entropy equal to $\log s$ ([MeS, Theorem II.8.1]).

Definition 15.8. We say a tent map is periodic if the orbit of the turning point is periodic. We call the tent map non-recurrent if the turning point (1/2) is non-recurrent.

The following is [MeS, Theorem II.8.1], due to Parry and to Milnor and Thurston. Note that for unimodal maps, showing uniqueness of the semi-conjugacy is relatively straightforward.

Fact 15.9. *Let $f : I \rightarrow I$ be a continuous unimodal map with positive entropy which fixes the boundary. Then there exists a unique continuous semi-conjugacy h_f to T_s , where $\log s$ is the (topological) entropy of f .*

Lemma 15.10. *Let $f \in \mathcal{F}$ have entropy $\log s > 0$ and let h_f be the conjugacy to T_s . If f has a periodic attractor then T_s is periodic. If T_s is non-recurrent, the semi-conjugacy h_f is a conjugacy and all periodic points of f are hyperbolic repelling and, if also $s > \sqrt{2}$, f is non-renormalisable.*

Proof. Since f has non-positive Schwarzian derivative, the turning point of f is contained in the immediate basin of a periodic attractor. The connected component of the immediate basin containing the critical point gets mapped by h_f to the turning point of T_s , which is therefore periodic, showing the first statement.

If T_s is non-recurrent then c is non-recurrent, so f has no periodic attractors and all periodic points are hyperbolic repelling, again by non-positive Schwarzian derivative. If $s > \sqrt{2}$ then T_s is non-renormalisable. Thus either f is non-renormalisable or f contains a restrictive interval containing c which is collapsed by h_f . In the latter case the (periodic) restrictive interval is mapped to the turning point, contradicting non-recurrence of T_s . Therefore f is non-renormalisable. Let us show (still in the case $s > \sqrt{2}$) that h_f is a conjugacy. By [MeS, Proposition III.4.3], we only need to show that f has no wandering intervals. Let W be a neighbourhood of c compactly contained in a larger neighbourhood W' disjoint from the post-critical orbit. If $f^k(x) \in W$ then $|Df^k(x)| > C$ for some constant C which depends only on the relative lengths of W and the connected components of $W' \setminus W$, using non-positive Schwarzian derivative. Thus no wandering interval can accumulate in W , as it cannot shrink indefinitely in size. On the other hand, any wandering interval must accumulate on c , for example by [MeS, Theorem IV.7.1]. In particular, f cannot have a wandering interval. This completes the case $s > \sqrt{2}$.

If $s \leq \sqrt{2}$, there is a least $k \geq 1$ such that $s' := 2^k \log s > 1/2$. The tent map T_s is then k times renormalisable of type 2. The k th renormalisation of T_s is the map $T_{s'}$. Pulling back, f is also k times renormalisable of type 2, with k th renormalisation f' say. But h_f now conjugates f' to $T_{s'}$, since $s' > \sqrt{2}$. It follows again that f and T_s are conjugate. \square

Corollary 15.11. *If $f \in \mathcal{F}$ has entropy $\log s > \log 2/2$ and the turning point of T_s is non-recurrent, then f is a non-renormalisable Misiurewicz map.*

Corollary 15.12. *If $f \in \mathcal{F}$ has entropy $\log s$ and T_s is non-recurrent, and if $g \in \mathcal{F}$, then f and g are conjugate if and only if the entropy of g also equals $\log s$.*

Let us continue with some further facts, deducible from or contained in [MeS, Chapter II], to which we also refer the reader for definitions and conventions. Since tent maps have no periodic attractors or wandering intervals, two tent maps with the same *kneading invariants* are conjugate and thus have the same entropy. Since the entropy is just logarithm of the slope, the maps themselves coincide. In particular, T_s and $T_{s'}$ have the same kneading invariants if and only if $s = s'$. With a little more work, one can then show the following: parameters s for which T_s is periodic are dense in $(1, 2]$; parameters s for which T_s is non-recurrent are dense in $(1, 2]$. Writing $\nu(s)$ for the kneading invariant of T_s , ν is a strictly monotone function. Moreover, at non-periodic parameters s , the kneading invariant $\nu(\cdot)$ is continuous (in general

it is not continuous at periodic parameters). Fixing T_s , the itinerary of a point x is denoted $\underline{i}_s(x)$, so $\nu(s) = \underline{i}_s(1/2)$. The itinerary map \underline{i}_s is also monotone. It is continuous at points whose orbits do not contain $1/2$. The itinerary of x is a sequence, for $n \geq 0$, whose n th element specifies whether $T_s^n(x) - 1/2$ is positive, negative or zero (that is, on which side of the turning point $T_s^n(x)$ lies). Denote by $\underline{i}_s^N(x)$ the itinerary sequence truncated to its first $N + 1$ elements.

Lemma 15.13. *Let $\sqrt{2} < s \leq 2$ and $p \neq 0$ be a periodic point for T_s with orbit $\text{Orb}(p)$ not containing the turning point. Then for all $\varepsilon > 0$ and $\delta \neq 0$ ($\delta < 0$ if $s = 2$), there is a parameter $r \in [s, s + \delta]$ for which T_r has a periodic turning point with period N say, and*

$$\#\{k \in \{1, \dots, N\} : T_r^k(1/2) \notin B(\text{Orb}(p), \varepsilon)\} < \varepsilon N.$$

Proof. We can suppose T_s is non-recurrent because non-recurrent parameters are dense in $(1, 2]$. By a simple perturbation argument, we can also assume that p is not in the orbit of the turning point. T_s is expanding on a neighbourhood of $\text{Orb}(p)$, so if $\varepsilon > 0$ is sufficiently small,

$$B(\text{Orb}(p), \varepsilon) \subset T_s(B(\text{Orb}(p), \varepsilon)).$$

There exist $q = q_{n_0}$ with $|q - p| < \varepsilon/2$ and some $n_0 \geq 1$ such that $T_s^{n_0}(q) = 1/2$. Therefore, there exists a sequence of points $q_n \in B(\text{Orb}(p), \varepsilon/2)$, $n \geq n_0$, for which $T_s(q_{n+1}) = q_n$ and $T_s^n(q_n) = 1/2$.

We can assume δ is small enough that there is a continuation $r \mapsto q_n^r$ for each n and all $r \in [s, s + \delta]$ satisfying $q_n^r \in B(\text{Orb}(p), \varepsilon)$, $T_r(q_{n+1}^r) = q_n^r$ for $n \geq n_0$, and $T_r^{n_0}(q_{n_0}^r) = 1/2$.

Let W denote the connected component of $I \setminus \text{Orb}(1/2)$. By transitivity, there are symmetric points $a, a_* \in W$, one on each side of $1/2$, and $l \geq 1$ such that $T_s^l(a) = T_s^l(a_*) = p$ and such that the orbits of a, a_* avoid the neighbourhood $V := (a, a_*)$ of $1/2$. We claim that there is a sequence of points y_j converging to $T_s(1/2)$ in $(T_s(1/2), T_{s+\delta}(1/2))$ and numbers k_j such that $T_s^{k_j}(y_j) = p$ and $\text{Orb}(y_j) \cap V = \emptyset$. Indeed, let η_j be small, positive or negative, and consider the images of the interval $(T_s(1/2), T_s(1/2) + \eta_j)$. Since T_s is expanding away from the turning point, and the orbit of $T_s(1/2)$ avoids W , the images grow until there is a first one which contains either a or a_* or both. Pull back the point to get a point y_j inside $(T_s(1/2), T_s(1/2) + \eta_j)$. The claim follows.

Combining these statements, we deduce that there are some (slightly smaller) neighbourhood V of $1/2$ and, given any increasing sequence of numbers $n_j > n_0$, a sequence of points x_j (close to the y_j) such that $T_s^{k_j}(x_j) = q_{n_j}$ and $T_s^l(x_j) \notin V$ for all $l < k_j + n_j$, while $T_s^{k_j+n_j}(x_j) = 1/2$. Since the k_j were already defined, the sequence of numbers n_j can be chosen with $n_j > k_j^2$, say, so $n_j/k_j \rightarrow \infty$. Provided the n_j are chosen large enough, the $\underline{i}_s(x_j)$ converge to $\underline{i}_s(T_s(1/2))$ in $(\underline{i}_s(T_s(1/2)), \underline{i}_{s+\delta}(T_{s+\delta}(1/2)))$.

Therefore, since the dynamics outside of V persists under small perturbation, there exist $\gamma \in (0, \delta)$, $N \geq 1$ and x_r for $r \in [s, s + \gamma]$ such that

- $r \mapsto x_r$ is continuous;
- $T_r(x_r) \neq 1/2$ for all $l < N$ and $r \in [s, s + \gamma]$.
- $T_r^N(x_r) = 1/2$;
- $\underline{i}_s^N(x_s) = \underline{i}_r^N(x_r) \in [\underline{i}_s^N(T_s(1/2)), \underline{i}_{s+\gamma}^N(T_{s+\gamma}(1/2))]$;
- $\#\{k \leq N : T_r^k(x_r) \notin B(\text{Orb}(p), \varepsilon)\} < \varepsilon N$.

In particular, for some $r \in [s, s + \gamma]$, $\underline{i}_r^N(T_r(1/2)) = \underline{i}_r^N(x_r)$, so $T_r(1/2) = x_r$ and $T_r^{N+1}(1/2) = 1/2$. \square

Lemma 15.14. *Let $\sqrt{2} < s < 2$ and let T_s be periodic with periodic turning point of period k . Let $b = s/(1+s)$ be the orientation-reversing fixed point of T_s . There are $N > 0$, a such that $T_s^N(a) = b$ and T_s^k is monotone and orientation-preserving on $[a, 1/2]$. For all r in some neighbourhood of s , the point a has a continuation a_r such that $T_r^N(a_r) = r/(1+r)$.*

There exist $s_n \rightarrow s$ such that, for all large n , the sequence

$$1/2, T_n^k(1/2), T_n^{2k}(1/2), \dots, T_n^{nk}(1/2) = a_n$$

is monotone, where (abusing notation) T_n denotes the tent map with slope s_n and a_n the corresponding continuation of the point a .

Proof. Fix a symmetric neighbourhood U of $1/2$ on which the slope of T_s^k only changes at $1/2$. Since T_s is non-renormalisable, there are pre-images of b in U . Let a be one such pre-image, and a_* its symmetric point, so for some $N > 0$, $T_s^N(a) = T_s^N(a_*) = b$. We can choose a such that $T_s^j(a) \notin (a, a_*)$ for $1 \leq j < N$. Then T_s^N is monotone on $(a, 1/2)$ and on $(1/2, a_*)$. Thus on one of those intervals, T_s^N is orientation-preserving.

The second statement is obvious.

The final statement follows from strict monotonicity of the kneading invariant for tent maps (and corresponds to the critical periodic orbit no longer existing, so the graph of T_n^k no longer quite touches the diagonal near $1/2$). \square

15.4. Proof of Theorem 1.30. Consider a continuous one-parameter family of maps $f_t \in \mathcal{F}$ (recall \mathcal{F} is endowed with the C^1 topology) where $t \in [0, \theta)$, for some $\theta > 0$. Let $\log s(t)$ denote the entropy of f_t . Suppose that f_0 is non-renormalisable and has all periodic points repelling, and that $t \mapsto s(t)$ is not locally-constant at $t = 0$. Non-renormalisability implies $s(0) > \sqrt{2}$. Taking θ smaller if necessary, we can assume that $\sqrt{2} < s(t) \leq 2$ for all $t \in [0, \theta)$.

Let $p \notin \partial I$ be a periodic point of f_0 with orbit $\text{Orb}(p)$ under iteration by f_0 . Fix $\varepsilon > 0$. We shall show that there exists $t \in (0, \theta)$ for which f_t is a non-renormalisable Misiurewicz map and for which the corresponding acip μ_t satisfies $\mu_t(B(\text{Orb}(p)), 2\varepsilon) > 1 - 2\varepsilon$, which suffices [note that approximate equidistribution along the orbit follows when ε is small].

Denote by h_t the semi-conjugacy map from f_t to $T_{s(t)}$. The following lemma says that if you are close enough to the corresponding tent map periodic orbit $h_0(\text{Orb}(p))$, then the pullback by h_t for small enough t is contained in an ε -neighbourhood of $\text{Orb}(p)$.

Lemma 15.15. *There exists $\gamma > 0$ such that, for all $t \in [0, \gamma]$,*

$$h_t(B(\text{Orb}(p), \varepsilon)) \supset B(h_0(\text{Orb}(p)), \gamma).$$

Proof. The point p is accumulated on both sides by repelling periodic points (note that h_0 is a conjugacy), and these move continuously as t varies, as do their images by h_t . Choose repelling periodic points q on each side of each point of $\text{Orb}(p)$ and choose γ' sufficiently small that the continuations q_t of these points for $t \in [0, 2\gamma')$ lie within $B(\text{Orb}(p), \varepsilon)$ but $h_t(q_t) \notin h_0(\text{Orb}(p))$. For each point $x \in h_0(\text{Orb}(p))$, there is a $\gamma_x > 0$ such that $h_t(q_t) \notin B(x, \gamma_x)$ for any q_t and any $t \in [0, \gamma']$. Take γ to be the minimum over $x \in h_0(\text{Orb}(p))$ of γ_x and of γ' . \square

Lemma 15.16. *Let γ be given by Lemma 15.15.*

There exists $s \in (s(0), s(\gamma))$ for which T_s is periodic with turning point of period k say, and

$$\#\{i \in \{1, \dots, k\} : T_s^i(1/2) \notin B(h_0(\text{Orb}(p)), \gamma)\} < \varepsilon k.$$

Proof. By Lemma 15.13. □

Let γ, s be given by Lemmas 15.15 and 15.16. Let the sequence $s_n \rightarrow s$ be given by Lemma 15.14. For large n , $s_n \in (s(0), s(\gamma))$. Thus there are maps $g_n \in \{f_t : t \in [0, \gamma)\}$ with entropy $\log s_n$ and there is a subsequence of the g_n converging to some map $g \in \mathcal{F}$ with entropy $\log s$.

It is straightforward to check that g has an orientation-preserving parabolic periodic point α , of period k say, which gets mapped by h_g to the turning point of T_s . We have from Lemma 15.16 that

$$\#\{i \in \{1, \dots, k\} : T_s^i(1/2) \notin B(h_0(\text{Orb}(p)), \gamma)\} < \varepsilon k,$$

so by Lemma 15.15

$$\#\{i \in \{1, \dots, k\} : g(\alpha) \notin B(\text{Orb}(p), \varepsilon)\} < \varepsilon k.$$

Applying Proposition 15.6, for sufficiently large n , the acip μ_n for g_n approximates the equidistribution on $\text{Orb}(\alpha)$ and so satisfies $\mu_n(B(\text{Orb}(p), 2\varepsilon)) > 1 - 2\varepsilon$. Since θ can be chosen arbitrarily small and $f_t := g_n$ is a non-renormalisable Misiurewicz map (by choice of s_n), this concludes the proof of Theorem 1.30.

16. POSITIVE ENTROPY DOES NOT IMPLY STATISTICAL STABILITY

In this section we show how lack of statistical stability may occur. The example will be explicit and shows (absolutely continuous) equilibrium states, with entropy bounded below, converging to a combination of an equilibrium state and a delta mass on a repelling fixed point.

We start with some general estimates for normalised quadratic maps. These estimates will transfer to any map affinely conjugate to the normalised maps. In the second subsection we define the sequence of maps and show instability, making use of the estimates in the first subsection.

Throughout this section, there will be topologically defined points (preimages of critical points and fixed points). Their dependence on parameters (of the maps considered) is frequently dropped from notation.

16.1. Quadratic maps which nearly have a fixed point. Up to rescaling and translation, all quadratic maps have the same form. We shall study the family $P_\kappa : x \mapsto x^2 + \kappa$, where κ is a real parameter. Each quadratic map is affinely conjugate (as a dynamical system) to some P_κ . For $\kappa = 1/4$, P_κ has a parabolic fixed point at $x_1 = 1/2$. For $\kappa > 1/4$, there is no fixed point and all points get mapped to the right. We are interested, for now, in how long it takes for points to pass some threshold.

Let $M > 10$ and consider the first entry time to $[M, \infty)$

$$e_{M, \kappa}(x) := \inf\{k \geq 0 : P_\kappa^k(x) \geq M\}.$$

For $M' \leq M^{2^j}$,

$$e_{M', \kappa} \leq e_{M, \kappa} + j. \tag{16.1}$$

Each function $e_{M,\kappa}$ is monotone-decreasing and piecewise-constant on $[0, M]$. For all $x > 0$, $j \geq 1$ and $\kappa > 1/4$,

$$\frac{dP_\kappa^j(x)}{d\kappa} > 0 \quad (16.2)$$

(visually, as one increases κ , the graph distances itself from the diagonal and images move to the right). Hence the bigger κ is, the less time it takes to enter $[M, \infty)$: for $1/4 < \kappa_0 < \kappa$, $e_{M,\kappa} \leq e_{M,\kappa_0}$.

Thanks to (16.2), given $V > 1$ and $n \geq 1$, there is a unique $\kappa(V, n) > 1/4$ such that $P_{\kappa(V,n)}^n(0) = V$. As a function of V , $\kappa(V, n)$ is strictly increasing. As n increases, $\kappa(V, n)$ decreases monotonically with limit $1/4$. Rather obviously, for $M \geq V$,

$$\int_{[0, 1/4]} e_{M,\kappa(V,n)} \geq (n-1)/4, \quad (16.3)$$

since $P(0) > 1/4$.

We wish to show that the first entry time to $[M, \infty)$ does not increase too fast as n increases, nor as V varies.

Lemma 16.1. *Given $V_0 > 100$, for all $M, M', V, V' \in (V_0, 100V_0)$ and $x \in [0, M]$,*

$$e_{M',\kappa(V',n+1)}(x) \leq e_{M,\kappa(V,n)}(x) + 3.$$

Proof. For now, set $M = V$ and let us compare $e_{V,\kappa(V,n)}$ and $e_{V,\kappa(V,n+1)}$. The discontinuities of the entry time functions lie along the orbits of 0, since $P_{\kappa(V,n)}^n(0) = V$. Let

$$\begin{aligned} p_j &:= P_{\kappa(V,n)}^j(0), \\ q_j &:= P_{\kappa(V,n+1)}^j(0). \end{aligned}$$

For $j > 0$, $q_j < p_j$. Since $p_n = q_{n+1} = M$ and $\kappa(V, n+1) < \kappa(V, n)$, $p_j < q_{j+1}$ for $j < n$. Since $e_{V,V,n} = n - j$ on $[p_j, p_{j+1})$ and

$$[p_j, p_{j+1}) \subset [q_j, q_{j+2})$$

we deduce

$$e_{V,\kappa(V,n)}(x) \leq e_{V,\kappa(V,n+1)}(x) \leq e_{V,\kappa(V,n)}(x) + 1. \quad (16.4)$$

Letting $V_0 > 100$ and $V, V' \in [V_0, 100V_0]$, (16.1) and (16.2) imply that

$$\begin{aligned} \kappa(V, n+1) &\leq \kappa(\sqrt{V}, n) \\ &\leq \kappa(V_0, n) \leq \kappa(V', n). \end{aligned}$$

Thus, $e_{V,\kappa(V',n)} \leq e_{V,\kappa(V,n+1)}$ for the same range of V, V' , and so

$$e_{V,\kappa(V',n)}(x) \leq e_{V,\kappa(V,n)}(x) + 1.$$

Replacing n by $n+1$ and using (16.4) yields

$$e_{V,\kappa(V',n+1)}(x) \leq e_{V,\kappa(V,n)}(x) + 2. \quad (16.5)$$

From (16.1) and (16.5) we obtain

$$\begin{aligned} e_{M',\kappa(V',n+1)}(x) &\leq e_{V,\kappa(V',n+1)}(x) + 1 \\ &\leq e_{V,\kappa(V,n)}(x) + 2 \leq e_{M,\kappa(V,n)}(x) + 3 \end{aligned}$$

for all $M, M', V, V' \in (V_0, 100V_0)$, as claimed. \square

We also need to estimate the distortion.

Lemma 16.2. *Let $\kappa > 1/4$. Let $y_0 \in [0, 1/4]$ and set $y_k := P_\kappa^k(y_0)$. Let k_0 be the minimal $k \geq 1$ for which $y_k \geq 1$. In particular, $DP(y_k) < 2$ if and only if $k < k_0$. First, if $k_0 \neq 1$ then*

$$\frac{DP_\kappa^{k_0-j}(z)}{DP_\kappa^{k_0-j}(z')} < 21 \quad (16.6)$$

for all $z, z' \in (y_{j-1}, y_j)$, for each $j = 2, 3, \dots, k_0$. Second,

$$y_{k_0} - y_{k_0-1} \geq 1/8. \quad (16.7)$$

Proof. We have $(y_1, y_{k_0-1}) \subset (1/4, 1)$, and on these intervals $1/2 < DP_\kappa < 2$ and $D^2P_\kappa = 2$. Hence $D \log DP_\kappa < 4$, and [noting $\log(Df(a)/Df(b)) = \int_b^a D \log Df$] we deduce the distortion bound

$$\frac{DP_\kappa^{k_0-j}(z)}{DP_\kappa^{k_0-j}(z')} \leq \exp\left(\int_{(1/4,1)} D \log DP_\kappa\right) < e^3 < 21,$$

for $j = 2, 3, \dots, k_0$ and all $z, z' \in (y_{j-1}, y_j)$.

To show the length estimate, since $y_{k_0} \geq 1$, we can assume $y_{k_0-1} > \sqrt{3}/2$, for otherwise the estimate holds trivially. Since $DP_\kappa(x) > 1$ for $x \geq 1/4 < \sqrt{3}/2$,

$$y_{k_0} - y_{k_0-1} > P_\kappa(\sqrt{3}/2) - \sqrt{3}/2 > 1 - \sqrt{3}/2 > 1/8.$$

□

16.2. Proof of Theorem 1.31. Theorem 1.31 avers the existence of a sequence $(f_k)_{k \geq 1} \in \mathcal{F}_{\text{NSD}}$ having decreasing critical relations with $f_k \rightarrow f_0$ as $k \rightarrow \infty$ and for which each f_k , $k \geq 0$ has an acip μ_k with entropy uniformly bounded away from 0, but with μ_k converging to a convex combination of μ_0 and a Dirac mass on a repelling fixed point.

Consider the family of maps

$$f_a : [0, 1] \rightarrow [0, 1],$$

with $a > 0$, defined by

$$f_a(x) = \begin{cases} 1 - 2x & \text{if } x \in [0, 1/2] \\ a(x - 1/2)(x - 1) + 1 & \text{if } x \in (1/2, 1]. \end{cases}$$

This map has two branches: one orientation-reversing, expanding branch on $[0, 1/2]$ and one unimodal branch on $(1/2, 1)$ with a critical point $c = 3/4$, see Figure 2. The map has fixed points at $1, 1/3$ and at $\alpha = \alpha(a)$, for some $\alpha \in (1/2, 1)$. Only parameters in $(0, 16)$ satisfy $f_a(c) \in (0, 1)$. For $a_0 = 32/3$, $f_{a_0}(c) = 1/3$, so the map is post-critically finite. Standard arguments (essentially a subset of those appearing at the end of this section) imply the existence of an acip for f_{a_0} ; the acip is the unique equilibrium state for the potential $-\log |Df_{a_0}|$ and has positive entropy.

The idea of the construction in Theorem 1.31 is the following. For a slightly bigger than a_0 , the critical orbit typically starts off spending a good amount of time close to $1/3$ and, after an even number of steps, escapes back to the interval $(1/2, 1]$. For certain parameters, the map f_a will have an acip which has a proportion of its mass very close to the repelling fixed point at $1/3$. We shall show this.

Remark 16.3. *If the orbit of c is disjoint from $\{0, 1/2, 1\}$, then f_a has the same critical relations as f_{a_0} .*

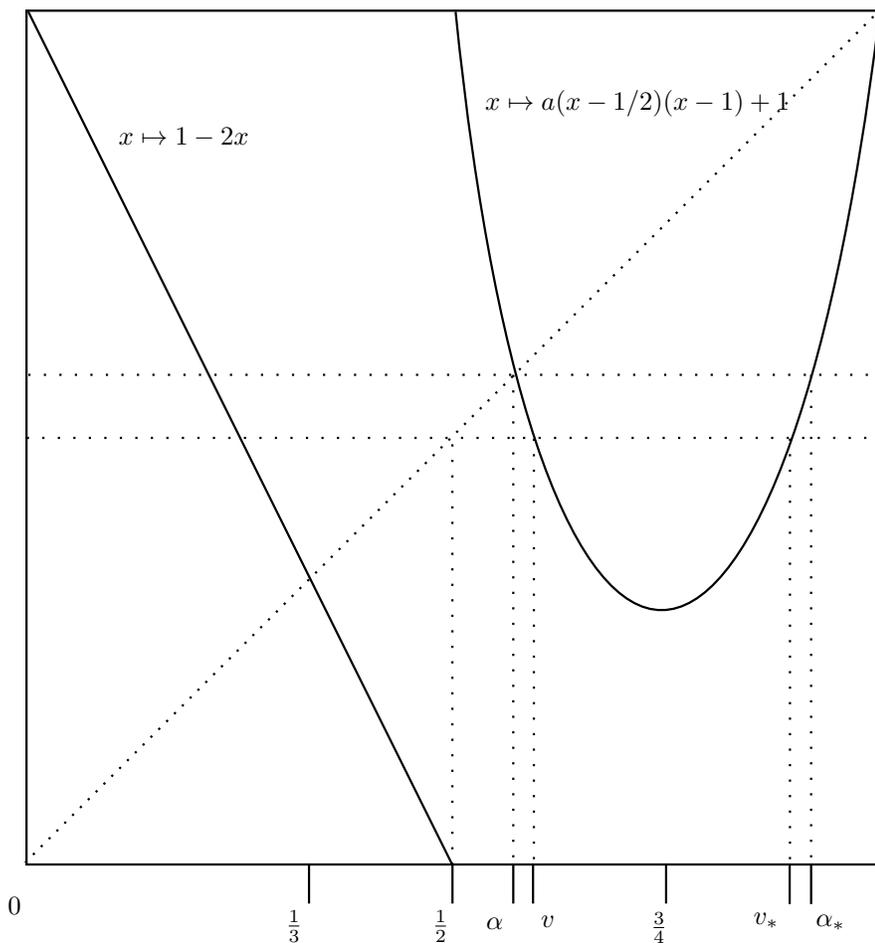


FIGURE 2. The map f_a for $a > a_0$

The following proposition implies Theorem 1.31.

Proposition 16.4. *Given $\beta > 0$, there exists a sequence $(a_k)_{k \geq 1}$ of parameters with $a_k \rightarrow a_0$ as $k \rightarrow \infty$ for which each f_{a_k} has an acip μ_k , the entropy of the μ_k is uniformly bounded below, and*

$$\lim_{k \rightarrow \infty} \mu_k = \frac{1}{1 + \beta} \mu_0 + \frac{\beta}{1 + \beta} \delta_{1/3},$$

where $\delta_{1/3}$ denotes the Dirac mass at the repelling fixed point $1/3$. Moreover the sequence $(f_{a_k})_{k \geq 0}$ has decreasing critical relations.

The proof of Proposition 16.4 occupies the remainder of this section. First we construct a first return map with one central branch (and lots of full branches). For a subset of parameters, taking higher iterates on the central branch will give a full-branched induced map. This will have an acip. These acips will converge to the acip for the induced map corresponding to a_0 . A careful selection of parameters will give control of the integral of the inducing time. If the integral of the inducing time on the (shrinking as $a \rightarrow a_0$) central branch converges to a non-zero constant, there is a good chance of proving the proposition.

For $a > a_0$, there is a point $\alpha_* > c$ (depending on a) with $f_a(\alpha_*) = f_a(\alpha) = \alpha$. There are points $v, v_* \in (\alpha, \alpha_*)$ (depending on a), with $v < v_*$, such that

$$f_a(v) = f_a(v_*) = 1/2.$$

For each even number $k \geq 2$, there exists a maximal interval A_k of parameters a for which $f_a^j(c) \in (0, 1/2)$ for $j = 1, \dots, k-1$, while $f_a^k(c) \in (1/2, 1)$. Moreover the map $\xi : a \rightarrow f_a^k(c)$ maps A_k continuously onto $(1/2, 1)$. By continuity, there is a subinterval A'_k of parameters in A_k mapped by ξ onto (α, α_*) . For a in A'_k , denote by $q < q_*$ the points closest to c with $f_a^k(q) = f_a^k(q_*) = \alpha$. As k is even, f_a^k has a local maximum at c and $f_a^k((q, q_*)) = (\alpha, c)$. The point v gets mapped by f_a to $1/2$, so $v, v_* \notin [q, q_*]$. In particular, (q, q_*) is compactly contained in (v, v_*) is compactly contained in (α, α_*) .

Inside the subinterval A'_k , there is a parameter $a(k, n)$ for which, writing $f_{k,n}$ for $f_{a(k,n)}$,

$$v = f_{k,n}^{kn}(c) < q < f_{k,n}^{k(n-1)}(c) < \dots < f_{k,n}^k(c) < c.$$

The graph of $f_{k,n}^k$ restricted to (q, q_*) lies below the diagonal. For $n \geq 2$ and k large, the graph very nearly touches the diagonal. Let us fix $f_{k,n}$ for now.

The first return map φ to (α, α_*) has a central branch defined on (q, q_*) and φ coincides with $f_{k,n}^k$ on (q, q_*) , see Figure 3. For $j = 1, \dots, n$,

$$\varphi^j(c) = f_{k,n}^{kj}(c).$$

Let us set

$$X := (\alpha, \alpha_*) \setminus [q, q_*].$$

Branches of φ in X are full, mapping diffeomorphically onto (α, α_*) [the proof of this is straightforward, noting that $\text{Orb}(\partial X) \cap X = \emptyset$]. Points in $[q, q_*]$ eventually leave $[q, q_*]$, and the first time they do, they enter $[\alpha, q]$. Denote by ψ the first entry map to X . Restricted to X it is just the identity map, while on (q, q_*) it consists of iterates of φ . Let us define the map

$$\Psi = \Psi_{k,n} := \varphi \circ \psi.$$

Lemma 16.5. *This map Ψ is full-branched and the following distortion bound holds for some constant C (independent of k, n):*

$$\left| \frac{D\Psi^n(x)}{D\Psi^n(y)} \right| \leq C |\Psi^n(x) - \Psi^n(y)| \quad (16.8)$$

for all x, y in the same branch of Ψ^n .

Proof. Since $\psi(c) = v$ and $f_{k,n}(v) = 1/2$, a fixed point at a positive (and bounded away from 0 for the parameters under consideration) distance from $[\alpha, \alpha_*]$, Ψ has no critical points (in the sense zero-derivative). Branches of ψ in (q, q_*) are mapped (by ψ) diffeomorphically onto (α, q) or onto (α, v) . Neither α nor v nor q is in any branch of φ . It follows that branches of Ψ in (q, q_*) are full.

Noting that the orbit of c never enters $(1/2, \alpha]$ nor $[\alpha_*, 1)$, each branch of Ψ extends (coinciding with the corresponding iterate of f) to map diffeomorphically onto $(1/2, 1)$. Moreover, the extensions have non-positive Schwarzian derivative. Therefore there is a uniform distortion bound for all iterates of Ψ , independently of k . \square

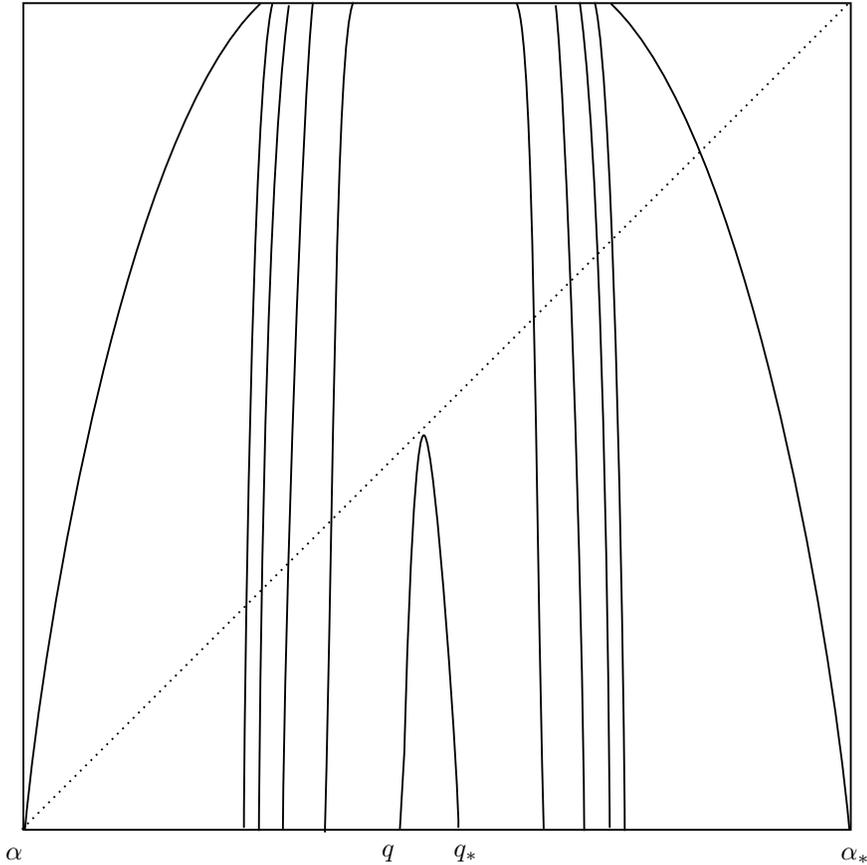


FIGURE 3. The first return map φ to (α, α_*) has one (almost parabolic) central branch (q, q_*) and infinitely many diffeomorphic branches.

On each branch of Ψ , Ψ coincides with some iterate of $f_{k,n}$. Thus there is a function $\tau = \tau_{k,n}$ defined on the domain of Ψ , constant on each branch of Ψ , for which

$$\Psi = f_{k,n}^\tau.$$

Estimating the size of $\tau^{-1}(j)$, for $j \in \mathbb{N}$, will give us information about the acip for $f_{k,n}$, namely its distribution, its Lyapunov exponent and (therefore) its entropy. Outside (q, q_*) , the estimates for τ which follow are independent of k, n .

Let E_j denote the set of points whose first entry time to $[\alpha, \alpha_*]$ is j . We shall now obtain uniform exponential bounds on the size of E_j .

Lemma 16.6. $m(E_j) \leq e^{-j/4}$.

Proof. Calculation gives $f_{a_0}(5/8) = 1/2$, so for $a > a_0$, $v < 5/8$. Therefore $\alpha < 5/8$ and $\alpha_* > 7/8$, so

$$m(E_0) = m((\alpha, \alpha_*)) \in (1/4, 1/2)$$

and

$$m(E_1) = 2m((\alpha_*, 1)) < 2^{-2}.$$

By convexity of $f_{k,n}$ on $(1/2, 1]$ and since $f_{k,n}(3/4) \leq 1/3$ and $f_{k,n}(1) = 1$, the graph of $f_{k,n}$ restricted to $(3/4, 1)$ lies under the line of slope $8/3 > 2$ passing through the repelling fixed point $(1, 1)$. Moreover, $f((\alpha_*, 1)) = (\alpha, 1)$. Therefore, for $j \geq 1$,

$$m(E_j \cap (\alpha_*, 1)) < 2^{-j-2}.$$

By symmetry,

$$m(E_j \cap (1/2, 1)) < 2^{-j-1}.$$

On $(0, 1/2)$, the set of points with first entry time to $(1/2, 1)$ equal to i has measure 2^{-i-1} , for $i \geq 1$, and this set is an interval mapped by $f_{k,n}^i$ onto $(1/2, 1)$. We deduce, for $j \geq 1$,

$$\begin{aligned} m(E_j \cap (0, 1/2)) &= \sum_{i=1}^j 2^{-i} m(E_{j-i} \cap (1/2, 1)) \\ &< \sum_{i=1}^j 2^{-i} 2^{-j+i-1} \\ &= j 2^{-j-1}. \end{aligned}$$

say. Hence $m(E_j) < (j+1)2^{-j-1}$, so

$$m(E_j) < e^{-j/4},$$

say. □

The distortion of $f_{k,n}$ on $(1/2, 1)$ is unbounded, due to the critical point. However, it is quadratic, which gives a certain amount of control. Denote by R_j the set of points in $[\alpha, \alpha_*]$ whose first return time to $[\alpha, \alpha_*]$ is equal to j .

Lemma 16.7. $m(R_j) \leq e^{-j/8}$.

Proof. Given a measurable set $X \subset (0, 1)$, since f_a is quadratic on $(1/2, 1)$,

$$m((1/2, 1) \cap f_a^{-1}(X)) \leq \sqrt{\frac{m(X)}{a}} \leq \sqrt{m(X)}/3,$$

provided $a \geq 32/3$. From the preceding lemma, we obtain

$$m(R_j) \leq \sqrt{m(E_{j-1})}/3 < e^{-\frac{j-1}{8}}/3 < e^{-j/8},$$

as required. □

Remark 16.8. *In the limit, Ψ converges (in some canonical sense) to the first return map for f_{a_0} to $[\alpha, \alpha_*]$. As the estimates of Lemma 16.7 are uniform, they also hold for the return map for f_{a_0} and the integral of the first return time, $\sum_j j m(R_j)$, is bounded.*

Returning to $f_{k,n}$, we can denote by $e_{k,n}$ the first entry time to $X = [\alpha, \alpha_*] \setminus [q, q_*]$. It is a multiple of k on (q, q_*) . We can decompose τ on (q, q_*) as

$$\tau = e_{k,n} + \tau \circ \psi.$$

We shall estimate these terms separately.

Lemma 16.9. $\lim_{k \rightarrow \infty} \sup_{n \geq 2} \int_{(q, q_*)} \tau \circ \psi dx = 0$.

Proof. The central branch is just

$$x \mapsto -2^k a(x - 3/4)^2 + 3/4 - \delta \quad (16.9)$$

for some small $\delta = \delta(k, n) > 0$. There is a sequence of points $q = q_0 < q_1 < q_2 < \dots < q_{n-1} < c$ for which $\varphi^l(q_l) = q$. Define the intervals $J_l := (q_l, q_{l+1})$ for $l \leq n-2$ and $J_{n-1} := (q_{n-1}, c)$. Each interval J_l , $l \geq 1$, gets mapped by φ onto J_{l-1} . The interval J_0 gets mapped onto (α, q) . On J_0 we have

$$m(\varphi^{-1}(R_j) \cap J_0) \leq \sqrt{\frac{m(R_j)}{2^k a}}.$$

Let $l_0 \geq 0$ be the maximal l for which $D\varphi(q_l) > 2$. For $1 \leq l < l_0$, on the interval J_l the derivative is at least 2. Hence

$$m(\psi^{-1}(R_j) \cap J_l) \leq 2^{-l} \sqrt{\frac{m(R_j)}{2^k a}}. \quad (16.10)$$

Going one step back, we obtain on J_{l_0} (taking a poor bound for better legibility),

$$m(\psi^{-1}(R_j) \cap J_{l_0}) \leq \left(\frac{m(R_j)}{2^k a} \right)^{1/4}.$$

Consider now the affine conjugacy A such that

$$\varphi = A^{-1} \circ P \circ A,$$

where P is the quadratic map $P : y \mapsto y^2 + \kappa$, and $\kappa > 1/4$ depends on δ and thus on k, n . One can calculate

$$A : x \mapsto -2^k a(x - 3/4) \quad (16.11)$$

(recalling $a = a(k, n)$). Since A is affine,

$$D\varphi(x) = DP(Ax). \quad (16.12)$$

Writing $y_0 := Aq_{n-1}$, we can apply Lemma 16.2. With the notation of that lemma, $k_0 = n - 1 - l_0$, while $y_{k_0-r} = Aq_{l_0+r}$. Thus $(y_{k_0-r-1}, y_{k_0-r}) = AJ_{l_0+r}$. For $k_0 - r = 2, 3, \dots, k_0$ or, equivalently, for $r = 0, 1, \dots, n - 3 - l_0$, by (16.6) and (16.12), for $x, x' \in J_{l_0+r}$,

$$\frac{D\varphi^r(x)}{D\varphi^r(x')} < 21.$$

Hence

$$m(\psi^{-1}(R_j) \cap J_{l_0+r}) \leq 21 \frac{|J_{l_0+r}|}{|J_{l_0}|} \left(\frac{m(R_j)}{2^k a} \right)^{1/4}.$$

Summing over intervals J_{l_0+r} , we obtain

$$m(\psi^{-1}(R_j) \cap (q_{l_0}, q_{n-2})) \leq 21 \frac{|q_{n-2} - q_{l_0}|}{|J_{l_0}|} \left(\frac{m(R_j)}{2^k a} \right)^{1/4}.$$

By (16.7), since A is affine, we have $|q_{n-2} - q_{l_0}| < 8|J_{l_0}|$, so

$$m(\psi^{-1}(R_j) \cap (q_{l_0}, q_{n-2})) \leq 168 \left(\frac{m(R_j)}{2^k a} \right)^{1/4}. \quad (16.13)$$

We have now dealt with (q_0, q_{n-2}) ; the intervals J_{n-2}, J_{n-1} remain to be treated. Pulling back (16.13) (once for J_{n-2} and once more for J_{n-1}), we obtain, for large k ,

$$m(\psi^{-1}(R_j) \cap (J_{n-2} \cup J_{n-1})) \leq \left(\frac{m(R_j)}{2^k a} \right)^{1/16}. \quad (16.14)$$

Combining the estimates $a > 1$, (16.10) (summed), (16.13) and (16.14) gives

$$\begin{aligned} m(\psi^{-1}(R_j) \cap (q, c)) &\leq 2\sqrt{\frac{m(R_j)}{2^k a}} + 168 \left(\frac{m(R_j)}{2^k a}\right)^{1/4} + \left(\frac{m(R_j)}{2^k a}\right)^{1/16} \\ &\leq 2^{10-k/16} m(R_j)^{1/16}. \end{aligned}$$

Note that there is no n -dependence in the bound.

From Lemma 16.7, $m(R_j) \leq \exp(-j/8)$. Hence

$$m(\psi^{-1}(R_j) \cap (q, c)) \leq 2^{10-k/16} \exp(-j/128).$$

Therefore

$$\int_{(q, q_*)} \tau \circ \psi \, dx = 2 \sum_j j m(\psi^{-1}(R_j) \cap (q, c)) \leq 2^{11-k/16} \sum_j j \exp(-j/128).$$

This upper bound has no n -dependence and tends to zero as $k \rightarrow \infty$, as required. \square

Lemma 16.10. *There is a constant $C_0 > 1$ such that, for all k, n ,*

$$\frac{1}{C_0} \sqrt{\frac{1}{2^k a}} \leq |q_* - q| \leq C_0 \sqrt{\frac{1}{2^k a}} \quad (16.15)$$

Proof. The image of φ on $[q, q_*]$ contains α and v , and $\inf_{k,n} |v - \alpha| > 0$. The range of φ has length bounded by $1/2$. From (16.9), the result then follows. \square

Lemma 16.11. *The first entry time $e_{k,n}$ to $X = [\alpha, \alpha_*] \setminus [q, q_*]$ has the following properties.*

•

$$\lim_{k \rightarrow \infty} \int_{(q, q_*)} e_{k,2} \, dx = 0; \quad (16.16)$$

•

$$\lim_{n \rightarrow \infty} \int_{(q, q_*)} e_{k,n} \, dx = \infty; \quad (16.17)$$

•

$$\int_{(q, q_*)} e_{k,n+1} \, dx \leq \int_{(q, q_*)} e_{k,n} \, dx + 4k|q_* - q|; \quad (16.18)$$

• For $x \in (q, q_*)$ and $j = 1, 2, \dots, k - s$,

$$\lim_{k \rightarrow \infty} \inf_{n \geq 1} \frac{\#\{j \leq e_{k,n}(x) : f_{k,n}^j(x) \in B(1/3, 1/k)\}}{e_{k,n}(x)} = 1. \quad (16.19)$$

Proof. The entry time satisfies $e_{k,n} \leq nk$, while (16.15) bounds the length of the interval (q, q_*) . This implies (16.16).

Recall the affine map A of (16.11) which conjugates the central branch of $\varphi = \varphi_{k,n}$ with a normalised quadratic $P : y \mapsto y^2 + \kappa$, with κ depending on k, n . [The map A also depends slightly on k, n .] By (16.15),

$$\frac{1}{2C_0} \sqrt{2^k a} \leq Aq \leq C_0 \sqrt{2^k a}.$$

Set $M_{k,n} := P(Aq) = A\alpha$ and $V_{k,n} := Av$. Then $P^n(0) = V_{k,n}$ (in particular, $\kappa = \kappa(V_{k,n}, n)$) and the first entry time at Ax to $[M_{k,n}, \infty)$ is $1 + \frac{e_{k,n}(x)}{k}$. From (16.3), we deduce (16.17).

Recall that $v \leq 5/8$, so $|\alpha - v| < |v - 3/4|$, so

$$M_{k,n}/2 \leq V_{k,n} \leq M_{k,n}.$$

Moreover, for fixed k ,

$$V_{k,2}/2 \leq V_{k,n} \leq 2V_{k,2}.$$

Setting $V_0 := 10V_{k,2}$, we can apply Lemma 16.1 with $M = M_{k,n}$, $V = V_{k,n}$, $M' = M_{k,n+1}$, $V' = V_{k,n+1}$. Via the conjugacy, this gives

$$e_{k,n+1} \leq e_{k,n}(x) + 3k.$$

This almost gives (16.18), we just replace 3 by 4 as the interval (q, q_*) depends on k, n and will vary a little for $n \geq 2$ (the entry time equals k on the non-common part).

For $x \in (q, q_*)$ and $j = 1, 2, \dots, k - s$, $f_{k,n}^j(x) \in B(1/3, 2^{-s})$. This implies (16.19). \square

To proceed, we need to recall some results concerning full-branched expanding maps applied to our situation.

Definition 16.12. *Let us call a collection of maps $(F_k)_{k \geq 0}$ a convergent Markov system if the following holds, denoting by \mathcal{D}_k the domain of F_k :*

- *There is an open interval I with $\mathcal{D}_k \subset I$ for all k , and $m(I \setminus \mathcal{D}_k) = 0$;*
- *each \mathcal{D}_k is a countable union of open intervals;*
- *for each branch J of F_k , $F_k : J \rightarrow I$ is a (surjective) C^1 diffeomorphism;*
- *for some K , for all $n, k \geq 0$, the distortion estimate*

$$\left| \frac{DF_k^n(x)}{DF_k^n(y)} \right| \leq K |F_k^n(x) - F_k^n(y)|$$

holds for all x, y in the same branch of F_k^n ;

- *the maps F_k converge uniformly to F_0 on compact subsets of \mathcal{D}_0 as $k \rightarrow \infty$.*

Then the following results are well-known, we provide only a sketch of their proof.

Fact 16.13. *Each F_k has an acip ν_k with Lipschitz density $\rho_k = \frac{d\nu_k}{dm}$ on I . The ρ_k are uniformly Lipschitz, bounded and bounded away from 0. The densities converge: $\rho_k \rightarrow \rho_0$ (uniformly) as $k \rightarrow \infty$. If τ_k is a continuous integer-valued function on \mathcal{D}_k , τ_k converges to τ_0 uniformly on compact subsets of \mathcal{D}_0 , and if there exists $C > 0$ such that, for all k ,*

$$m(\{\tau_k^{-1}([n, \infty))\}) < Cn^{-3},$$

then for each $\gamma > 0$, there exists a compact set $\Lambda \subset \mathcal{D}_0$ satisfying

$$\int_{\mathcal{D}_k \setminus \Lambda} \tau_k d\nu_k < \gamma \tag{16.20}$$

for $k = 0$ and for all large k , while $\tau_k = \tau_0$ on Λ for all large k .

Proof. The first two statements are proven in [MeS, Theorem V.2.2] (at least for one map, but the estimates are uniform, depending only on the distortion estimate). We now sketch the proof of the final statement, using ideas similar to those earlier in the paper. One can find, thanks to the distortion bounds, a uniform N, λ for which $|DF_k^N| > \lambda > 1$. It follows that return maps to small intervals have exponential tails: if P is a subinterval of I then the set of points with first return time $\geq j$ has measure $\leq C(|P|)e^{-j\delta}$, where $C(\cdot) > 0$ can be taken monotone increasing and only depending on the length of P , and both $C(\cdot)$ and $\delta > 0$ are independent of k . If n

is large enough and P is a branch of F_0^n , then P is tiny and by the uniform bounds on the densities, the densities ρ_k are approximately constant on P . By Kac' Lemma,

$$\int_P r_k d\rho_k = 1, \quad (16.21)$$

where r_k denotes the return time to P for F_k . Given small $\varepsilon > 0$, let Λ be a finite union of intervals compactly contained in some Λ_1 compactly contained in the domain of the return map to P under F_0 , and choose Λ (and Λ_1) such that

$$\int_{P \setminus \Lambda} r_k d\rho_k < \varepsilon. \quad (16.22)$$

This we can do by the uniform exponential tails estimate. Note that r_0 is bounded on Λ , and that Λ does not depend on k . For k large enough, r_k coincides with r_0 on Λ (by uniform convergence of F_k on Λ_1). But then, by (16.21) and (16.22),

$$\left| 1 - \int_{\Lambda} r_0 d\rho_k \right| = \left| 1 - \int_{\Lambda} r_k d\rho_k \right| \leq \varepsilon$$

and similarly

$$\left| 1 - \int_{\Lambda} r_0 d\rho_0 \right| \leq \varepsilon.$$

Since the ρ_k are each approximately constant (*a priori* varying a lot with k) on P and bounded away from zero and infinity, we deduce that $\rho_0 \approx \rho_k$ for large k , or $\rho_k \rightarrow \rho_0$ as $k \rightarrow \infty$.

It remains to prove the statements concerning τ_k . As a continuous integer-valued function, τ_k is constant on each connected component of \mathcal{D}_k . Given any compact subset of $\Lambda \subset \mathcal{D}_0$, for k large enough $\tau_k \equiv \tau_0$ on Λ . The tail estimate implies that, given $\gamma > 0$, there exists $\varepsilon > 0$ such that on any set Y of Lebesgue measure $\varepsilon > 0$,

$$\int_Y \tau_k d\nu_k < \gamma.$$

In particular, for Λ large enough that $\mathcal{D}_k \setminus \Lambda$ has measure at most ε for all large k (and for $k = 0$),

$$\int_{\mathcal{D}_k \setminus \Lambda} \tau_k d\nu_k < \gamma$$

for k large and for $k = 0$. showing (16.20). \square

Returning to our collection of maps $f_{k,n}$ with their induced maps $\Psi_{k,n}$ and inducing times τ_k , note that (q, q_*) tends to the point $\{3/4\}$ as $k \rightarrow \infty$ (independently of n). Moreover the maps $\Psi_{k,n}$ have a uniform distortion bound (16.8). Let us denote by Ψ_0 the first return map to $[\alpha, \alpha_*]$ under f_{a_0} , and by τ_0 its return time. Then for any sequence $(n_k)_k$, setting $\Psi_k := \Psi_{k,n_k}$, the sequence $(\Psi_k)_{k \geq 0}$ is a convergent Markov system.

In particular, the acips ν_k for Ψ_k converge the acip ν_0 for Ψ_0 . By Remark 16.8,

$$S_0 := \int \tau_0 d\nu_0 < \infty.$$

On the other hand, if we choose n_k carefully, the integral of the inducing times will converge, but not to the corresponding integral S_0 for Ψ_0 . Presently we shall choose n_k .

Denote by $S_{k,n}$ the integral of the inducing time for $\Psi_{k,n}$, for $k \geq 0$:

$$S_{k,n} := \int_{[\alpha, \alpha_*]} \tau_k d\nu_k.$$

Recall $X = X_{k,n} = [\alpha, \alpha_*] \setminus [q, q_*]$. It will be useful to define

$$S_{k,n}^X := \int_X \tau_k d\nu_k$$

and

$$S_{k,n}^\delta := \int_{(q,q_*)} \tau_k d\nu_k$$

so $S_{k,n}^X + S_{k,n}^\delta = S_{k,n}$.

Given $\beta > 0$, for $k \geq 1$, let $n(k) \geq 2$ be the minimal n such that $S_{k,n}^\delta > \beta S_0$. For large k ,

$$S_{k,n}^\delta \approx \int_{(q,q_*)} e_{k,n} d\nu_k$$

by Lemma 16.9. Now (16.15) and the first three statements of Lemma 16.11 together imply that $n(k)$ exists and $S_{k,n(k)}^\delta \approx \beta S_0$. In particular,

$$\lim_{k \rightarrow \infty} S_{k,n(k)}^\delta = \beta S_0.$$

Henceforth, we fix the sequence $n_k := n(k)$. We can drop the n -dependence from the notation in the terms considered, and write S_k for $S_{k,n}$, etc. We denote the parameter $a(k, n_k)$ by a_k . As $k \rightarrow \infty$, S_k^X converges to S_0 . This follows from convergence of the Lipschitz densities ρ_k (Fact 16.13), convergence of the maps Ψ_k and the exponential tail estimate for the inducing times τ_k (Lemma 16.7, as $\tau_k(x) = j$ on $R_j \cap X$).

Now let us look at the spread measures. For $k \geq 0$, let ν_k^j denote the restriction of ν to $\tau_k^{-1}(j)$ and set

$$\mu_k := \frac{1}{S_k} \sum_{j \geq 1} \sum_{i=0}^{j-1} (f_{a_k}^i)_* \nu_k^j.$$

Then μ_k is an ergodic acip for each $k \geq 0$. We wish to show that μ_k does not converge to μ_0 as $k \rightarrow \infty$. With this aim, for $k \geq 1$, let us restrict ν_k to X and then spread:

$$\eta_k^X := \frac{1}{S_k} \sum_{j \geq 1} \sum_{i=0}^{j-1} (f_{a_k}^i)_* (\nu_k^j)|_X.$$

We do the same for ν_k restricted to (q, q_*) :

$$\eta_k^\delta := \frac{1}{S_k} \sum_{j \geq 1} \sum_{i=0}^{j-1} (f_{a_k}^i)_* (\nu_k^j)|_{(q,q_*)}.$$

Then $\mu_k = \eta_k^X + \eta_k^\delta$ and

$$\eta_k^X([0, 1]) = \frac{S_k^X}{S_k}.$$

Because tails on X are exponential and Ψ_k restricted to X converges to Ψ_0 , one can use Fact 16.13 and in particular (16.20) to show that η_k^X converges to

$$\lim_{k \rightarrow \infty} \frac{S_k^X}{S_k} \mu_0 = \frac{1}{1 + \beta} \mu_0.$$

So the question is what happens to the limit of η_k^δ ? Well, let J_j denote the set of $x \in (q, q_*)$ with $e_{k,n_k}(x) = j$. The difference between η_k^δ and the measure

$$\hat{\eta}_k^\delta := \frac{1}{S_k} \sum_{j \geq 1} \sum_{i=0}^j (f_{a_k}^i)_*(\nu_k^j)|_{J_j}$$

is tiny, by Lemma 16.9. In particular, the limits of η_k^δ and $\hat{\eta}_k^\delta$ coincide (assuming they exist).

By (16.19),

$$\lim_{k \rightarrow \infty} \hat{\eta}_k^\delta([0, 1] \setminus B(1/3, 1/k)) = 0.$$

Therefore $\hat{\eta}_k^\delta$ and η_k^δ converge to an atomic measure supported on the point $1/3$ with mass given by

$$\lim_{k \rightarrow \infty} \frac{S_k^\delta}{S_k} = \frac{\beta}{1 + \beta}.$$

It remains to bound from below the entropies of μ_k . The Lyapunov exponents $\int_{[\alpha, \alpha_*]} \log |D\Psi_k| d\nu_k$ are uniformly bounded below away from 0. By Pesin's formula, the entropies $h(\nu_k)$ are also uniformly bounded below away from 0. The integrals of the inducing time converge to $S_0(1 + \beta)$. By Abramov's formula, the entropies $h(\mu_k)$ are also bounded away from 0.

That the critical relations are decreasing follows from Remark 16.3 and Definition 1.13.

This concludes the proof of Proposition 16.4 which, we recall, implies Theorem 1.31.

REFERENCES

- [A] J. Aaronson, *An introduction to infinite ergodic theory*, volume 50 of Mathematical Surveys and Monographs. American Mathematical Society, Providence, RI, 1997.
- [Ab1] L.M. Abramov, *The entropy of a derived automorphism*, Dokl. Akad. Nauk SSSR **128** (1959) 647–650.
- [ALM] L. Alsedà, J. Llibre, M. Misiurewicz, *Combinatorial dynamics and entropy in dimension one*, volume 5 of Advanced Series in Nonlinear Dynamics. World Scientific Publishing Co. Inc., River Edge, NJ, second edition, 2000.
- [Al] J.F. Alves, *Strong statistical stability of non-uniformly expanding maps*, Nonlinearity **17** (2004) 1193–1215.
- [ACF] J.F. Alves, M.P. Carvalho, J.M. Freitas, *Statistical Stability for Hénon maps of the Benedicks-Carleson type*, Ann. Inst. H. Poincaré Anal. Non Linéaire, **27** (2010) 595–637.
- [AS] J.F. Alves, M. Soufi, *Statistical stability of Lorenz attractors*, Fund. Math. **224** (2014) 219–231.
- [AV] J.F. Alves, M. Viana, *Statistical stability for robust classes of maps with non-uniform expansion*, Ergodic Theory Dynam. Systems **22** (2002) 1–32.
- [Ar] V. Araújo, *Semicontinuity of entropy, existence of equilibrium states and continuity of physical measures*, Discrete Contin. Dyn. Syst. **17** (2007) 371–386.
- [Ba1] V. Baladi, *On the susceptibility function of piecewise expanding interval maps*, Comm. Math. Phys. **275** (2007) 839–859.
- [Ba2] V. Baladi, *Linear response despite critical points*, Nonlinearity **21** (2008) T81–T90.
- [BBS] V. Baladi, M. Benedicks, D. Schnellmann, *Whitney-Holder continuity of the SRB measure for transversal families of smooth unimodal maps*, Invent. Math. **201** (2015) 773–844.
- [BaS1] V. Baladi, D. Smania, *Linear response formula for piecewise expanding unimodal maps*, Nonlinearity **21** (2008) 677–711.
- [BaS2] V. Baladi, D. Smania, *Analyticity of the SRB measure for holomorphic families of quadratic-like Collet-Eckmann maps*, Proc. Amer. Math. Soc. **137** (2009) 1431–1437.
- [BC] M. Benedicks, L. Carleson, *On iterations of $1 - ax^2$ on $(-1, 1)$* , Ann. Math. **122** (1985) 1–25.
- [BM] M. Benedicks, M. Misiurewicz, *Absolutely continuous invariant measures for maps with flat tops*, Inst. Hautes Études Sci. Publ. Math. **69** (1989) 203–213.
- [BL] A.M. Blokh, M.Y. Lyubich, *Ergodicity of transitive unimodal transformations of the interval*, Ukrain. Mat. Zh. **41** (1989) 985–988.

- [Bo] R. Bowen, *Symbolic dynamics for hyperbolic systems, Equilibrium states and the ergodic theory of Anosov diffeomorphisms*, Lecture Notes in Math. **470** Springer, Berlin 1975.
- [B] H. Bruin, *Induced maps, Markov extensions and invariant measures in one-dimensional dynamics*, Commun. Math. Phys. **168** (1995) 571–580.
- [BK] H. Bruin, G. Keller, *Equilibrium states for S -unimodal maps*, Ergodic Theory Dynam. Systems **18** (1998) 765–789.
- [BKNS] H. Bruin, G. Keller, T. Nowicki, S. van Strien, *Wild Cantor attractors exist*, Ann. of Math. **143** (1996) 97–130.
- [BT] H. Bruin, M. Todd, *Equilibrium states for interval maps: the potential $-t \log |Df|$* , Ann. Sci. École Norm. Sup. (4) **42** (2009) 559–600.
- [Bu] J. Buzzi, *On entropy-expanding maps*, Preprint CMLS, 2000.
- [C] S. Cedervall, *Invariant Measures and Correlation Decay for S -multimodal Interval Maps*, PhD thesis, Imperial College, London, 2006.
- [CE] P. Collet, J.-P. Eckmann, *Positive Liapunov exponents and absolute continuity for maps of the interval*, Ergodic Theory Dynam. Systems **3** (1983) 13–46.
- [Col] J.F. Colombeau, *Differential Calculus and Holomorphy*, North Holland Math. Studies 64, 1982.
- [CR] D. Coronel, J. Rivera-Letelier, *Low-temperature phase transitions in the quadratic family*, Adv. Math. **248** (2013) 453–494.
- [Co] G. Contreras, *Regularity of topological and metric entropy of hyperbolic flows*, Math. Z. **210** (1992) 97–111.
- [D1] N. Dobbs, *Renormalisation-induced phase transitions for unimodal maps*, Comm. Math. Phys. **286** (2009) 377–387.
- [D2] N. Dobbs, *On cusps and flat tops*, Ann. Inst. Fourier **64** (2014) 571–605.
- [EM] P. Eslami, M. Misiurewicz, *Singular limits of absolutely continuous invariant measures for families of transitive maps* J. Difference Equ. Appl. **18** (2012) 739–750.
- [F] J. Freitas, *Continuity of SRB measure and entropy for Benedicks-Carleson quadratic maps*, Nonlinearity **18** (2005) 831–854.
- [FT] J. Freitas, M. Todd, *Statistical stability of equilibrium states for interval maps*, Nonlinearity **22** (2009) 259–281.
- [GPR] K. Gelfert, F. Przytycki, M. Rams, *On the Lyapunov spectrum for rational maps*, Math. Ann. **348** (2010) 965–1004.
- [GS] J. Graczyk, G. Świątek, *Generic hyperbolicity in the logistic family*, Ann. of Math. (2) **146** (1997) 1–52.
- [H1] F. Hofbauer, *On intrinsic ergodicity of piecewise monotonic transformations with positive entropy*, Israel J. Math., **34** (1979) 213–237.
- [H2] F. Hofbauer, *Piecewise invertible dynamical systems*, Probab. Theory Relat. Fields **72** (1986) 359–386.
- [HR] F. Hofbauer, P. Raith, *The Hausdorff dimension of an ergodic invariant measure for a piecewise monotonic map of the interval*, Canad. Math. Bull. **35** (1992) 84–98.
- [HomY] A.J. Homburg, T. Young, *Intermittency and Jakobson’s theorem near saddle-node bifurcations*, Discrete Contin. Dyn. Syst. **17** (2007) 21–58.
- [IT1] G. Iommi, M. Todd, *Natural equilibrium states for multimodal maps*, Comm. Math. Phys. **300** (2010) 65–94.
- [IT2] G. Iommi, M. Todd, *Dimension theory for multimodal maps*, Ann. Henri Poincaré **12** (2011) 591–620.
- [J] M.V. Jakobson, *Absolutely continuous invariant measures for one-parameter families of one-dimensional maps*, Comm. Math. Phys. **81** (1981) 39–88.
- [KKPW] A. Katok, G. Knieper, M. Pollicott, H. Weiss, *Differentiability and analyticity of topological entropy for Anosov and geodesic flows*, Invent. Math. **98** (1989) 581–597.
- [K1] G. Keller, *Stochastic stability in some chaotic dynamical systems*, Monatsh. Math. **94** (1982) 313–333.
- [K2] G. Keller, *Lifting measures to Markov extensions*, Monatsh. Math. **108** (1989) 183–200.
- [K3] G. Keller, *Equilibrium states in ergodic theory*, London Mathematical Society Student Texts, 42. Cambridge University Press, Cambridge, 1998.
- [Ko] O.S. Kozlovski, *Getting rid of the negative Schwarzian derivative condition* Ann. of Math. (2) **152** (200) 743–762.
- [L] F. Ledrappier, *Some properties of absolutely continuous invariant measures on an interval*, Ergodic Theory Dynam. Systems **1** (1981) 77–93.
- [LGBPE] Z. Li, P. Góra, A. Boyarsky, H. Proppe, P. Eslami, *Family of piecewise expanding maps having singular measure as a limit of ACIMs*, Ergodic Theory Dynam. Systems **33** (2013) 158–167.
- [LLF] A. Libchaber, C. Laroche, S. Fauve, *Period doubling cascade in mercury, a quantitative measurement*, J. Physique Lett. **43** (1982) 211–216.
- [Ly] M. Lyubich, *Dynamics of quadratic polynomials I-II*, Acta Math. **178** (1997) 185–297.

- [MU] R.D. Mauldin, M. Urbański, *Graph directed Markov systems*, volume 148 of Cambridge Tracts in Mathematics. Cambridge University Press, Cambridge, 2003. Geometry and dynamics of limit sets.
- [MN] I. Melbourne, M. Nicol, *Almost sure invariance principle for nonuniformly hyperbolic systems*, *Comm. Math. Phys.* **260** (2005) 131–146.
- [MeS] W. de Melo, S. van Strien, *One dimensional dynamics*, *Ergebnisse Series* **25**, Springer–Verlag, 1993.
- [M1] M. Misiurewicz, *Structure of mappings of an interval with zero entropy*, *Inst. Hautes Études Sci. Publ. Math.* **53** (1981) 5–16.
- [M2] M. Misiurewicz, *Possible jumps of entropy for interval maps*, *Qual. Theory Dyn. Syst.* **2** (2001) 289–306.
- [MS] M. Misiurewicz, S. Shlyachkov, *Entropy of piecewise continuous interval maps*, *European Conference on Iteration Theory (Batschuns, 1989)*, 239–245, World Sci. Publ., River Edge, NJ, 1991.
- [PSe] Y. Pesin, S. Senti, *Equilibrium measures for maps with inducing schemes*, *J. Mod. Dyn.* **2** (2008) 397–427.
- [Pi] V. Pinheiro, *Expanding measures*, *Ann. Inst. H. Poincaré Anal. Non Linéaire* **28** (2011) 889–939.
- [P] F. Przytycki, *Lyapunov characteristic exponents are nonnegative*, *Proc. Amer. Math. Soc.* **119** (1993) 309–317.
- [PR1] F. Przytycki, J. Rivera-Letelier, *Geometric pressure for multimodal maps of the interval*, To appear in *Mem. Amer. Math. Soc.* (arxiv:1405.2443).
- [PR2] F. Przytycki, J. Rivera-Letelier, *Nice inducing schemes and the thermodynamics of rational maps*, *Comm. Math. Phys.* **301** (2011) 661–707.
- [PRS] F. Przytycki, J. Rivera-Letelier, S. Smirnov, *Equality of pressures for rational functions*, *Ergodic Theory Dynam. Systems* **24** (2004) 891–914.
- [PU] F. Przytycki, M. Urbański, *Conformal fractals: ergodic theory methods*, volume 371 of London Mathematical Society Lecture Note Series. Cambridge University Press, Cambridge, 2010.
- [Ra] P. Raith, *Stability of the maximal measure for piecewise monotonic interval maps*, *Ergodic Theory Dynam. Systems* **17** (1997) 1419–1436.
- [R] D.A. Rand, *The singularity spectrum $f(\alpha)$ for cookie-cutters*, *Ergodic Theory Dynam. Systems* **9** (1989) 527–541.
- [Ri] J. Rivera-Letelier, *Asymptotic expansion of smooth interval maps*, Preprint (arXiv:1204.3071).
- [Ru1] D. Ruelle, *A measure associated with Axiom A attractors*, *Am. J. Math.* **98** (1976) 619–654.
- [Ru2] D. Ruelle, *Repellers for real analytic maps*, *Ergodic Theory Dynam. Systems* **2** (1982) 99–107.
- [Ru3] D. Ruelle, *Differentiation of SRB states*, *Comm. Math. Phys.* **187** (1997) 227–241.
- [Ru4] D. Ruelle, *A review of linear response theory for general differentiable dynamical systems*, *Nonlinearity* **22** (2009) 855–870.
- [Ru5] D. Ruelle, *Structure and f -dependence of the A.C.I.M. for a unimodal map f of Misiurewicz type*, *Comm. Math. Phys.* **287** (2009) 1039–1070.
- [RS] M. Rychlik, E. Sorets, *Regularity and other properties of absolutely continuous invariant measures for the quadratic family*, *Comm. Math. Phys.* **150** (1992) 217–236.
- [Si] Y. Sinai, *Gibbs measures in ergodic theory*, (Russian) *Uspehi Mat. Nauk* **27** (1972) 21–64. English translation: *Russian Math. Surveys* **27** (1972) 21–69.
- [Su] D. Sullivan, *Bounds, quadratic differentials, and renormalization conjectures*. In *American Mathematical Society centennial publications, Vol. II* (Providence, RI, 1988), pages 417–466. Amer. Math. Soc., Providence, RI, 1992.
- [Th] H. Thunberg, *Unfolding of chaotic unimodal maps and the parameter dependence of natural measures*, *Nonlinearity* **14** (2001) 323–337.
- [Ts1] M. Tsujii, *Positive Lyapunov exponents in families of one-dimensional dynamical systems*, *Invent. Math.* **111** (1993) 113–137.
- [Ts2] M. Tsujii, *On continuity of Bowen-Ruelle-Sinai measures in families of one dimensional maps*, *Comm. Math. Phys.* **177** (1996) 1–11.
- [VV] P. Varandas, M. Viana, *Existence, uniqueness and stability of equilibrium states for non-uniformly expanding maps*, *Ann. Inst. H. Poincaré Anal. Non Linéaire* **27** (2010) 555–593.
- [V] C.H. Vázquez, *Statistical stability for diffeomorphisms with dominated splitting*, *Ergodic Theory Dynam. Systems* **27** (2007) 253–283.
- [Y] L.-S. Young, *Statistical properties of dynamical systems with some hyperbolicity*, *Ann. of Math. (2)* **147** (1998) 585–650.
- [Y] L.-S. Young, *Recurrence times and rates of mixing*, *Israel J. Math.* **110** (1999) 153–188.
- [Z] R. Zweimuller, *Invariant measures for general(ized) induced transformations*, *Proc. Amer. Math. Soc.* **133** (2005) 2283–2295.

SCHOOL OF MATHEMATICS AND STATISTICS, UNIVERSITY COLLEGE DUBLIN, BELFIELD, DUBLIN 14, IRELAND.

E-mail address: `neil.dobbs@ucd.ie`

URL: <http://www.maths.ucd.ie/~ndobbs/>

MATHEMATICAL INSTITUTE, UNIVERSITY OF ST ANDREWS, NORTH HAUGH, ST ANDREWS, FIFE, KY16 9SS,
SCOTLAND

E-mail address: `mjt20@st-andrews.ac.uk`

URL: <http://www.mcs.st-and.ac.uk/~miket/>