Ergod. Th. & Dynam. Sys., (2023), **43**, 3150–3192
doi:10.1017/etds.2022.60

Quenched and annealed equilibrium states for random Ruelle expanding maps and applications

MANUEL STADLBAUER†, PAULO VARANDAS‡§ and XUAN ZHANG¶

† *Instituto de Matemática, Universidade Federal do Rio de Janeiro, Rio de Janeiro 21941-909, RJ, Brazil*
(e-mail: manuel@im.ufrj.br)

‡ *CMUP, Faculdade de Ciências, Universidade do Porto, Porto 4169-007, Portugal*

§ *Departamento de Matemática, Universidade Federal da Bahia, Salvador 40170-115, BA, Brazil*
(e-mail: paulo.varandas@ufba.br)

¶ *Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo 05508-090, SP, Brazil*
(e-mail: xuan@ime.usp.br)

(*Received* 7 *January* 2022 *and accepted in revised form* 25 *July* 2022)

Abstract. We find generalized conformal measures and equilibrium states for random dynamics generated by Ruelle expanding maps, under which the dynamics exhibits exponential decay of correlations. This extends results by Baladi [Correlation spectrum of quenched and annealed equilibrium states for random expanding maps. *Comm. Math. Phys.* **186** (1997), 671–700] and Carvalho *et al* [Semigroup actions of expanding maps. *J. Stat. Phys.* **116**(1) (2017), 114–136], where the randomness is driven by an independent and identically distributed process and the phase space is assumed to be compact. We give applications in the context of weighted non-autonomous iterated function systems, free semigroup actions and introduce a boundary of equilibria for not necessarily free semigroup actions.

Key words: random dynamical systems, quenched and annealed equilibrium states, non-autonomous dynamical systems, semigroup actions
2020 Mathematics Subject Classification: 37A25, 37C85, 37D35 (Primary); 37H05, 47B80, 37A50 (Secondary)

Contents

1 Introduction 3151
2 Statement of the main results 3152

3	Semigroups of Ruelle expanding maps on non-compact spaces	3155
4	Quotients of Ruelle operators	3159
5	Contraction in the Wasserstein distance	3162
6	Conformal measures, quenched exponential decay and continuity	3166
7	Annealed exponential decay	3171
8	An almost sure invariance principle	3177
9	Applications	3181
Acknowledgements		3190
References		3190

1. *Introduction*

In this paper, we contribute to the thermodynamic formalism of sequential and random dynamical systems, whose notions we now recall. Given a compact metric space X, a probability space (Ω, P), a measurable map $\theta : \Omega \to \Omega$ and a family $(T_\omega)_{\omega\in\Omega}$ of maps acting on X, one is interested in describing typical points according to the random orbit

$$T_\omega^n := T_{\theta^{n-1}(\omega)} \circ \cdots \circ T_{\theta(\omega)} \circ T_\omega. \tag{1.1}$$

For each fixed $\omega \in \Omega$, the previous expression consists of the iteration of the sequential dynamical system $(T_n)_n$, with $T_n := T_{\theta^n\omega}$. The random transformation associated to the family $(T_\omega)_{\omega\in\Omega}$ and randomness $(\Omega, \theta, \mathrm{P})$ can be modelled by the skew-product

$$\begin{aligned} F : \Omega \times X &\to \Omega \times X \\ (\omega, x) &\mapsto (\theta(\omega), T_\omega(x)). \end{aligned}$$

The space of F-invariant probability measures whose marginal on Ω is given by P is non-empty and every such probability μ is characterized by the disintegration

$$d\mu(\omega, x) = d\mu_\omega(x)\, d\mathrm{P}(\omega), \tag{1.2}$$

where μ_ω are called the sample measures of μ. The previous expression encloses the information of the sequential dynamics arising from the random dynamical system. Indeed, a description of the dynamics as in equation (1.1) for P-typical points ω allows for the description of the probabilities μ_ω and the reconstruction of the whole random dynamics through equation (1.2). The previous formalism has proved to be very useful to code the dynamics of finitely generated semigroup actions, in which case one obtains a step skew-product F (see e.g. [**6**, **7**, **20**, **33**, **34**] and references therein).

In view of the previous discussion, it is natural that one of the central questions in the thermodynamic formalism for random dynamics is how to effectively construct conformal-like (and equilibrium state-like) measures, as it might allow one to establish, for example, limit laws or stability under perturbations. This goal has been attained in several variations of the setting above. If θ is an ergodic automorphism and the T_ω are expanding maps, then there are several known versions of a quenched Ruelle–Perron–Frobenius theorem, a line of research which was initiated by works of Bogenschütz–Gundlach and Kifer [**4**, **21**]. That is, the classical statement of the theorem holds for P-almost every sequence of transfer operators dual to (T_ω^n). By combining the result with a random version of the variational principle, this then gives rise to the notion of equilibrium states as well

https://doi.org/10.1017/etds.2022.60 Published online by Cambridge University Press

as their uniqueness (see [26] and references therein, or e.g. the recent contributions in [1]). In a purely topological context of fibred systems with Ruelle expanding fibres and a homeomorphism as factor, Denker, Gordin and Heinemann [11, 12] obtained a quenched version of Ruelle's theorem and a construction of relative equilibrium states. However, these questions have also been studied for arbitrary sequences of expanding maps on the unit interval [9, 19] or general non-autonomous dynamical systems (we refer the reader to [8, 18] and references therein).

Alternatively, the annealed setting approaches these notions in average with respect to P. If the base is an independent and identically distributed stochastic process, it was shown by Baladi [2] that the annealed equilibria are the averages of the quenched ones with respect to P. The restriction to independent and identically distributed processes in there is a consequence of the simple observation that the independence implies that taking averages with respect to P and the iterations of the quenched transfer operators commute.

A further, related approach to these questions is to consider the semigroup generated by the maps $\{T_\omega\}$. However, even though semigroups and random iterations of these maps are intrinsically different, the results in [6, 7] indicate that the associated thermodynamic formalism might bridge this gap and should give rise to an important field of applications.

A motivation for our work is the attempt to unify the above settings for the case of a finite family of distance expanding maps on Polish spaces. Starting from a technical result on geometric convergence of a family of quenched operators, we deduce two quenched versions of Ruelle's theorem and a description of the fluctuations of the quenched ergodic sums through a central limit theorem for the quenched setting. Moreover, in the random regime, these results imply geometric convergence of the averaged operators with respect to a ψ-mixing, non-invertible transformation θ in the base and a formula for the almost sure Hausdorff dimension of the limit sets of a random conformal iterated function system. Finally, it follows from these quenched results that one may identify a topological boundary of the semigroup with the set of quenched equilibrium states, and that this identification is Lipschitz continuous.

2. *Statement of the main results*

In what follows, we introduce the setting and state the main results of this paper. However, for the sake of simplicity, we postpone several technical definitions to the next sections. Throughout, we assume that (X, d) is a complete and separable metric space, and that $T_1, \ldots T_k : X \to X$ are continuous, surjective and Ruelle expanding maps (cf. Definition 3.2). Moreover, we always assume that the semigroup $\mathcal{S}$ generated by these maps is jointly topologically mixing and finitely aperiodic (cf. Definitions 3.3 and 3.4).

Moreover, as we are interested in thermodynamic quantities, we fix Hölder continuous functions $\varphi_1, \ldots, \varphi_k : X \to \mathbb{R}$ and define, for a finite word $v = i_1 \ldots i_n$,

$$T_v := T_{i_n} \circ \cdots \circ T_{i_1} \quad \text{and} \quad \varphi_v := \varphi_{i_1} + \varphi_{i_2} \circ T_{i_1} + \cdots + \varphi_{i_n} \circ T_{i_1 i_2 \ldots i_{n-1}}.$$

This then gives rise to a family of Ruelle operators $\{L_v\}$ and a further family of operators $\{\mathbb{P}_u^v\}$, defined by

$$L_v(f)(x) := \sum_{T_v(y)=x} e^{\varphi_v(y)} f(y), \quad \mathbb{P}_u^v(f) = \frac{L_v(f \cdot L_u(\mathbf{1}))}{L_{uv}(\mathbf{1})},$$

for f in a suitable function space and with $\mathbf{1}$ referring to the constant function of value 1. Moreover, to guarantee that $L_v(\mathbf{1})$ is well defined, we also assume that the functions φ_i are summable (cf. Definition 4.1). As it will turn out below, the analysis of this family of operators allows us to ignore the problem of the non-existence of invariant densities due to purely functorial reason and was, according to the authors' knowledge, first employed in [3].

The two main features of these quotients are that $\mathbb{P}_u^v(\mathbf{1}) = \mathbf{1}$ and that the iteration rule $\mathbb{P}_{uv}^w \circ \mathbb{P}_u^v = \mathbb{P}_u^{vw}$ holds. It follows from the first that the dual operators $\{(\mathbb{P}_u^v)^*\}$ act on the space of probability measures $\mathcal{M}_1(X)$, and from the second that it is possible to adapt methods for Markov operators as in [5, 17, 23, 31] to obtain geometric convergence. Our first principal result now establishes this kind of convergence. In here, $\overline{W}$ refers to the Wasserstein metric and $\overline{D}$ to the Hölder coefficient with respect to the equivalent metric d^* (cf. equation (5.1)). We refer the reader to §4 for the necessary definitions and notation.

THEOREM A. *Suppose the Ruelle expanding semigroup $\mathcal{S}$ is jointly topologically mixing and finitely aperiodic, and that every potential φ_i is α-Hölder and summable. Then there exist $k_0 \in \mathbb{N}$ and $s \in (0, 1)$ such that for all finite words u, v with length $|v| \geq k_0$ and $\nu_1, \nu_2 \in \mathcal{M}_1(X)$ and every Hölder continuous observable $f : X \to \mathbb{R}$ with $\overline{D}(f) < \infty$,*

$$\overline{W}(\mathbb{P}_u^{v*}(\nu_1), \mathbb{P}_u^{v*}(\nu_2)) \leq s^{|v|}\overline{W}(\nu_1, \nu_2),$$
$$\overline{D}(\mathbb{P}_u^v(f)) \leq s^{|v|}\overline{D}(f).$$

This theorem implies that for any infinite word $\omega = i_1 i_2 \ldots$ and measure $\nu \in \mathcal{M}_1(X)$, the limit

$$\mu_\omega := \lim_{l\to\infty} (\mathbb{P}_\emptyset^{i_1\ldots i_l})^*(\nu)$$

exists, is independent of ν and the speed of convergence is exponential. This means that, under some mild assumptions on the set of Ruelle expanding maps, any non-autonomous sequence of dynamics admits a probability measure that rules its dynamics and that this measure is a non-autonomous conformal measure in the following sense: there exists $\lambda_{u,\omega} > 0$ such that $L_u^*(\mu_\omega) = \lambda_{u,\omega}\mu_{u\omega}$ (see Proposition 6.1). Furthermore, for any left infinite word $\tilde{\omega} = \ldots i_{-2}i_{-1}$, the limit

$$\mu_{\tilde{\omega},\omega} := \lim_{l\to\infty} (\mathbb{P}_{i_{-l}\ldots i_{-1}}^{i_1\ldots i_l})^*(\nu)$$

exists, varies Hölder continuously with ω, is independent of ν, and the speed of convergence is exponential. As shown in Proposition 6.3, this measure is invariant in the non-autonomous setting, and if $\tilde{\omega}$ and ω are periodic extensions of the finite word w, that is, $\tilde{\omega} = \ldots ww$ and $\omega = ww\ldots$, then $\mu_{\tilde{\omega},\omega}$ is the unique equilibrium state of (T_w, φ_w) (cf. Proposition 6.5). In fact, the set of all measures $\{\mu_{\tilde{\omega},\omega}\}$, where $\tilde{\omega}, \omega$ run through all infinite words is the closure of these equilibrium states and can be used to define a compactification of the semigroup (Proposition 9.4).

A further application of Theorem A is related to an invariance principle as the contraction allows us to apply the general invariance principle in [10] and gives rise to

https://doi.org/10.1017/etds.2022.60 Published online by Cambridge University Press

the following result (for a similar result for continued fractions with restricted entries, see [32]). Here, $[\omega]_n$ stands for the initial n-word of an infinite word ω.

THEOREM B. *Suppose the finitely Ruelle expanding semigroup $\mathcal{S}$ is jointly topologically mixing and finitely aperiodic, and that every potential φ_i is α-Hölder and summable. Suppose $\omega \in \Sigma$, $f \in \mathcal{H}_\alpha$. Let $f_n = f - \int f \circ T_{[\omega]_n}\, d\mu_\omega$ for every $n \in \mathbb{N}_0$, and let $s_n^2 = \mathbb{E}_{\mu_\omega}(\sum_{k=0}^{n-1} f_k \circ T_{[\omega]_k})^2$ for $n \geq 1$ and assume that $\sum_n s_n^{-4} < \infty$. Then there exists a sequence (Z_n) of independent centred Gaussian random variables such that*

$$\sup_n \left| \sqrt{\textstyle\sum_{k=0}^{n-1} \mathbb{E}_{\mu_\omega} Z_k^2} - s_n \right| < \infty,$$

$$\sup_{0 \leq k \leq n-1} \left| \sum_{i=0}^{k} f_i \circ T_{[\omega]_i} - \sum_{i=0}^{k} Z_i \right| = o\left(\sqrt{s_n^2 \log \log s_n^2} \right) \textit{ almost surely.}$$

We then relate and apply these results to random dynamical systems, that is, we assume that the T_i are chosen with respect to a given probability measure ρ. So, it is sufficient to fix a measure ρ either on the shift spaces $\Sigma := \{1, \ldots, k\}^{\mathbb{N}}$ or $\Sigma_{\mathbb{Z}} := \{1, \ldots, k\}^{\mathbb{Z}}$ and consider the almost sure behaviour, referred to as *quenched*, and the behaviour in average, referred to as *annealed* behaviour. In this setting, Proposition 6.1 provides existence and exponential decay towards the quenched random conformal measure μ_ω, whereas the bilateral result in Proposition 6.3 implies the same statement for the quenched equilibrium state $\mu_{\tilde{\omega},\omega}$.

To relate these quenched results to their annealed counterparts, we consider in here as in [2] the annealed operators

$$\mathcal{A}_n := \sum_{|w|=n} \rho(\{\omega : [\omega]_n = w\}) L_w.$$

A fundamental problem of these operators is that, in general, $\mathcal{A}_{n+m} \neq \mathcal{A}_n \circ \mathcal{A}_m$, which makes it impossible to apply methods from spectral theory. However, if we assume that ρ is supported on a topologically mixing, one-sided subshift of finite type, it is possible to control the asymptotic behaviour of $\{\mathcal{A}_n\}$, which is our third main result. In here, θ refers to the one-sided shift map.

THEOREM C. *Suppose the Ruelle expanding semigroup $\mathcal{S}$ is jointly topologically mixing and finitely aperiodic, and that every potential φ_i is α-Hölder and summable. Moreover, suppose that ρ is supported on a topologically mixing, one-sided subshift of finite type and that $d\rho/d\rho \circ \theta$ is Hölder continuous. Then there exist $r \in (0, 1)$, a positive function $h \in \mathcal{H}_\alpha$ and $\beta > 0$ such that for all $f \in \mathcal{H}_\alpha$ and every large $n \geq 1$,*

$$\left| \frac{\mathcal{A}_n(f)(x)}{\beta^n h(x)} - \int f\, d\pi \right| \ll r^n (\overline{D}(f) + \|f\|_m).$$

Now assume that ρ is a Bernoulli measure, so that the maps T_i are chosen independently. Then, by independence, it follows that $\mathcal{A}_n = (\mathcal{A}_1)^n$. Hence, as an immediate corollary, one obtains that

https://doi.org/10.1017/etds.2022.60 Published online by Cambridge University Press

$$(\mathcal{A}_1)^n(hf)(x)/\beta^n h(x) \longrightarrow \int f(x)h(x)\,d\pi(x)$$

exponentially fast, which is a well-known version of Ruelle's operator theorem for independently chosen maps T_i (cf. Proposition 3.1 in [2]). As this is the key step for existence and uniqueness of the annealed equilibrium state (cf. Proposition 3.3 in [2]), one obtains Theorem 1 in [2] for independent and identically distributed Ruelle expanding maps as a corollary.

We now return to the general case of a one-sided subshift of finite type with exponential decay of correlations and now assume, in addition, that ρ is θ-invariant. In this setting, we obtain an annealed version of decay of correlations.

THEOREM D. *Suppose that the assumptions of Theorem C hold and that ρ is θ-invariant. Then there exist a probability measure $\tilde{\pi}$, $r \in (0, 1)$ and $k_1 \in \mathbb{N}$ such that*

$$\bigg| \int \sum_{|v|=n} \mathbf{1}_{[v]}(\omega) f(T_v(x)) g(x)\,d\mu_\omega(x)\,d\rho(\omega) - \int f\,d\tilde{\pi} \int g\,d\mu_\omega\,d\rho \bigg|$$
$$\leq r^n \int |f|\,d\mu_\omega\,d\rho\bigg(\overline{D}(g) + \int |g|\,d\mu_\omega\,d\rho\bigg)$$

for all $g \in \mathcal{H}_\alpha$ and $f : X \to \mathbb{R}$ integrable with respect to $d\mu_\omega(x)\,d\rho(\omega)$.

The latter reveals an unexpected connection between quenched and annealed dynamics. Indeed, it is noticeable that despite the fact that quenched and annealed random dynamical systems often measure different complexities of the dynamics (see e.g. [6, Proposition 8.3] for an explicit formula in the context of free semigroup actions), in Theorem D, we obtain an annealed decay of correlations with respect to a probability $d\mu_\omega\,d\rho$ obtained via quenched asymptotics. These results for both quenched and annealed dynamical systems will appear as Theorems 5.1, 7.3, 7.4 and 8.3 below. Moreover, the authors would like to point out that, according to their knowledge, Theorems C and D are the first annealed results for a dependent choice of the maps $\{T_i\}$. Finally, in §9, we discuss applications to non-autonomous conformal iterated function systems, the thermodynamic formalism of semigroup actions and a boundary construction through equilibrium states.

3. *Semigroups of Ruelle expanding maps on non-compact spaces*

We always assume that (X, d) is a complete and separable metric space and that $\mathcal{W}$ is a finite alphabet. For every $i \in \mathcal{W}$, let $T_i : X \to X$ be a continuous, surjective transformation and let $\mathcal{S}$ be the semigroup generated by $\{T_i\}_{i\in\mathcal{W}}$, that is,

$$\mathcal{S} = \{T_{i_k} \circ T_{i_{k-1}} \circ \cdots \circ T_{i_1} : k \in \mathbb{N},\ i_1, i_2, \ldots, i_k \in \mathcal{W}\}.$$

For every $k \in \mathbb{N}$ and every finite word $v = i_1 i_2 \ldots i_k \in \mathcal{W}^k$, set

$$T_v := T_{i_k} \circ \cdots \circ T_{i_1}.$$

Then each element of $\mathcal{S}$ is equal to T_v for some finite word v, but v might not be uniquely determined (e.g. if two generators T_a, T_b commute, then $T_{ab} = T_{ba}$). Observe that, with the usual concatenation of words, we have that $T_{vw} = T_w \circ T_v$ and, in particular, that the

https://doi.org/10.1017/etds.2022.60 Published online by Cambridge University Press

map from $\bigcup_{k\geq 1} \mathcal{W}^k \to \mathcal{S}$ given by $v \mapsto T_v$ is a semigroup anti-homomorphism, referred to as the coding of $\mathcal{S}$. This coding naturally defines a free semigroup action $\mathcal{S} \times X \to X$, $(T_v, x) \mapsto T_v(x)$ determined by $\mathcal{S}$.

For every finite word $v \in \mathcal{W}^k$, denote its length by $|v| = k$. For $x \in X$ and $A \subset X$, let $B_r(x) = \{y \in X : d(x, y) < r\}$ and $B_r(A) = \{y \in X : d(x, y) < r$ for some $x \in A\}$. For a finite word $v = i_1 \ldots i_k$, define dynamical distance

$$d_v(x, y) := \sup\{d(x, y),\ d(T_{i_1\ldots i_j}(x), T_{i_1\ldots i_j}(y)), 1 \leq j < |v|\}$$

and dynamical ball

$$B_r^v(x) := \{y \in X : d_v(x, y) < r\}.$$

Later, we will also consider infinite words. The transformations $T_i, i \in \mathcal{W}$ in this paper are always *Ruelle expanding* maps as introduced in [29]. However, here, we do not require that the base space is compact and, in particular, the set of preimages of a point might be countably infinite. Recall that this notion of expanding map is defined as follows.

Definition 3.1. T is said to be (a, λ)-*Ruelle expanding*, for some $a > 0$ and $\lambda \in (0, 1)$, if for any $x, y, \tilde{x} \in X$ with $d(x, y) < a$ and $T(\tilde{x}) = x$, there exists a unique $\tilde{y} \in X$ with $T(\tilde{y}) = y$ and $d(\tilde{x}, \tilde{y}) < a$, and such that this $\tilde{y}$ satisfies

$$d(\tilde{x}, \tilde{y}) \leq \lambda d(x, y).$$

Examples of Ruelle expanding maps include C^1 expanding maps on compact Riemannian manifolds, distance expanding maps on compact metric spaces and one-sided subshifts of countable type. In particular, our setting includes distance expanding maps on non-compact metric spaces. Observe that as we only consider a finite alphabet $\mathcal{W}$, we may choose the same parameters a and λ for all $T_i, i \in \mathcal{W}$.

Definition 3.2. The semigroup $\mathcal{S}$ generated by $\{T_i\}_{i\in\mathcal{W}}$ is said to be a (a, λ)-*Ruelle expanding semigroup* if every $T_i, i \in \mathcal{W}$ is (a, λ)-Ruelle expanding.

We extend to the semigroup $\mathcal{S}$ the notions of topological mixing and finite aperiodicity, which are usually defined for the iteration of a single map. They are known from graph directed Markov systems [25] or from the big images and preimages property for shift spaces [30].

Definition 3.3. $\mathcal{S}$ is said to be *jointly topologically mixing* if for all open sets $U, V \subset X$, there exists $m \in \mathbb{N}$ such that $T_w^{-1}(U) \cap V \neq \emptyset$ for all finite words w with $|w| \geq m$.

Definition 3.4. An (a, λ)-Ruelle expanding semigroup $\mathcal{S}$ is said to be *n-finitely aperiodic* (see Figure 1) if there exist $n \in \mathbb{N}$, a finite subset $K \subset X$ and $r > 0$ such that for all $x \in X$ and $w \in \mathcal{W}^n$, one can find $\xi, \eta \in K$ satisfying:

(1) there is $\xi^* \in T_w^{-1}(\xi)$ with $d_w(x, \xi^*) < a$;
(2) there is $x^* \in T_w^{-1}(x)$ with $d(x^*, \eta) < a$ and $d_w(x^*, \eta) < r$.

The first condition is modelled after the big image condition, the second after the big preimage condition.

https://doi.org/10.1017/etds.2022.60 Published online by Cambridge University Press

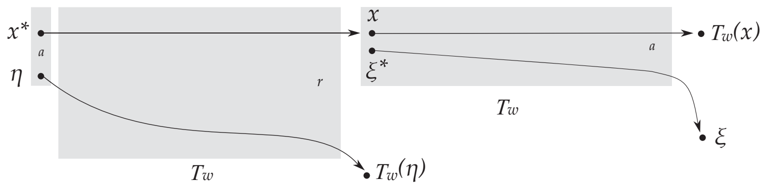

FIGURE 1. Finite aperiodicity.

Remark 3.1. Any Ruelle expanding semigroup defined on a compact space X is n-finitely aperiodic for every $n \in \mathbb{N}$, which can be seen by the following argument. Let K be a finite set such that $X \subset \bigcup_{z \in K} B_{a/2}(z)$ and let $r = \operatorname{diam}(X)$. Choose $\xi \in K \cap B_a(T_w(x))$, then the Ruelle expanding property assures the existence of ξ^* and hence condition (1). Choose any $x^* \in T_w^{-1}(x)$ and $\eta \in K \cap B_a(x^*)$, then condition (2) follows.

We now present two classes of examples of jointly topologically mixing and finitely aperiodic semigroups.

Example 3.2. Assume that (X, d) is a compact and pathwise-connected metric space such that there exists some $C > 0$ such that for any pair $(x, y) \in X$, there exists a rectifiable curve from x to y of length smaller than C. Furthermore, assume that $\{T_i\}_{i \in \mathcal{W}}$ is a finite family of Ruelle expanding maps on X.

PROPOSITION 3.3. *$\{T_i\}_{i \in \mathcal{W}}$ is jointly topologically mixing and finitely aperiodic.*

Proof. By Remark 3.1, it remains to show that the semigroup is jointly topologically mixing. To do so, we show that for any open set $U \subset X$, there exists $m \in \mathbb{N}$ such that $T_w(U) = X$ for all finite words w with $|w| \geq m$.

So assume that $x, y \in X$ are connected by a curve γ_0 of length $\ell(\gamma_0) \leq C$ and that $i \in \mathcal{W}$. By covering γ with finitely many open balls of radius a and by choosing for each of these open balls an inverse branch of T_i such that the inverse branches coincide in the overlapping regions of the covering, one obtains a new curve γ_1 such that $T_i(\gamma_1) = \gamma_0$. Furthermore, as T_i is a local homeomorphism whose inverse branches contract distances by λ, it follows that γ_1 is rectifiable and that $\ell(\gamma_1) \leq \lambda \ell(\gamma_0)$. It hence follows by iteration that for any w with $|w| = n$, there exists a curve γ_n with $T_w(\gamma_n) = \gamma_0$ and $\ell(\gamma_n) \leq C\lambda^n$.

So assume that U contains an open ball with centre z of radius r, that $r < C\lambda^n$, that $|w| = n$ and that $x \in X$. Then, for a curve γ_0 of length $\ell(\gamma_0) \leq C$ from $T_w(z)$ to $x \in X$, there exists a curve γ_n which starts in z such that $T_w(\gamma_n) = \gamma_0$ and $\ell(\gamma_n) \leq C\lambda^n < r$. Hence, the endpoint of γ_n is an element of U. As x is arbitrary, it follows that $T_w(U) = X$. $\square$

Example 3.4. We now construct a class of semigroups generated by a finite number of skew products over the same topological Markov chain and provide sufficient conditions for joint topological mixing and finite aperiodicity.

https://doi.org/10.1017/etds.2022.60 Published online by Cambridge University Press

To do so, we recall the notion of a topological Markov chain with the big images and preimages property. So assume that $A = (a_{ij})_{i,j\geq 0}$ is a matrix with values in $\{0, 1\}$ without rows or columns equal to 0. We then refer to

$$\Sigma := \{(x_i : i \in \mathbb{N} \cup \{0\}) : x_i \in \mathbb{N} \cup \{0\}, a_{x_i x_{i+1}} = 1 \text{ for all } i \geq 0\}$$

as a *topological Markov chain* with *transition matrix* A. Furthermore, we say that A is *aperiodic* if for any pair (i, j), there exists $n_0 \in \mathbb{N}$ such that the coordinate (i, j) of the nth power A^n is strictly positive for all $n > n_0$. Moreover, we say that Σ has the *big images and preimages property* if there exits a finite subset $L \subset \mathbb{N} \cup \{0\}$ such that for each $n \in \mathbb{N} \cup \{0\}$, there exist $k, l \in L$ such that $a_{kn} = 1$ and $a_{nl} = 1$. It is worth noting here that the non-triviality of rows and columns imply that Σ is non-compact with respect to the product of the discrete topology on $\mathbb{N} \cup \{0\}$. In combination with the big images and preimages property, this then implies that Σ is even locally non-compact.

We now show that the left shift $\sigma : \Sigma \to \Sigma$ is a topologically mixing 1-aperiodic Ruelle expanding map with respect to the metric $d_\sigma((x_i), (y_i)) := 2^{-\min\{i : x_i \neq y_i\}}$, which is compatible with the product topology on Σ. First, note that $d_\sigma(x, y) \leq 3/4$ implies that x and y share the same first coordinate. In particular, the restriction of σ on balls of radius 3/4 is a homeomorphism and expands distances by 2. That is, σ is $(\frac{3}{4}, \frac{1}{2})$-Ruelle expanding. Moreover, it follows from aperiodicity of A and finiteness of L that there exists m_0 such that for any pair (i, j) in L, $\sigma^{m_0}([i]) \subset [j]$, where $[a] \subset \Sigma$ refers to those elements in Σ, whose first coordinate is equal to a. Hence, it follows from big images and preimages that $\sigma^{m_0+2}([a]) = \Sigma$ for any $a \in \mathbb{N} \cup \{0\}$. This then implies that σ is topologically mixing. To see that σ is 1-aperiodic in the sense of Definition 3.4, it remains to choose for each $i \in L$ an element $x_i \in [i]$ and check that $\{x_i : i \in L\}$ satisfies the conditions of Definition 3.4.

Now fix (X, d) is as in Example 3.2, $\lambda \in (0, 1)$, $a > 0$ and a finite set $\mathcal{W}$. Furthermore, assume that the set of (a, λ)-Ruelle expanding maps on X is non-empty and that for any $w \in \mathcal{W}$, κ_w associates to each $\mathbb{N} \cup \{0\}$ a Ruelle expanding map, that is,

$$\kappa_w : \mathbb{N} \cup \{0\} \to \{T : X \to X \mid T \text{ is } (a, \lambda)\text{-Ruelle expanding}\}.$$

In particular, κ_w gives rise to the skew product

$$T_w : \Sigma \times X \to \Sigma \times X, \ \ ((x_i), y) \mapsto (\sigma((x_i)), T_{\kappa_w(x_0)}(y))$$

and the semigroup $\mathcal{S}$ generated by $\{T_w : w \in \mathcal{W}\}$. With respect to $d_{\mathcal{S}}((x, y), (\bar{x}, \bar{y})) := d_\sigma(x, \bar{x}) + d(y, \bar{y})$, one then obtains the following.

PROPOSITION 3.5. *$\mathcal{S}$ is jointly topologically mixing and 1-aperiodic.*

Proof. Assume without loss of generality that $a \leq 1/2$. Then, $d_{\mathcal{S}}((x, y), (\bar{x}, \bar{y})) := d_\sigma(x, \bar{x}) + d(y, \bar{y}) < a$ implies that the first coordinate of x and $\bar{x}$ coincide and that $d(y, \bar{y}) < a$. Hence, it follows that the restriction of T_w to a ball of radius a is a homeomorphism and that the inverse branches of T_w contract at least with rate $\max\{1/2, \lambda\}$. Now assume that U is open. Then there exist $k \in \mathbb{N}$, $x_0, \ldots x_k \in \mathbb{N} \cup \{0\}$ and $r > 0$ such that $[x_0, \ldots x_k] \times B_r(z) \subset U$, where $[x_0, \ldots x_k]$ refers to those elements in Σ starting with $x_0, \ldots x_k$ and $B_r(z)$ to the ball of radius r with centre z in X. It now follows from the above that $\sigma^{k+m_0+2}([x_0, \ldots x_k]) = \Sigma$ and from Example 3.2 that $T_w(B_r(z)) = X$ for any w with $C\lambda^{|w|} < r$. In particular, there exists n with $T_w(U) = \Sigma \times X$ for any $w \in \mathcal{W}^n$. In particular, $\mathcal{S}$ is jointly topologically mixing. The remaining statement that is the finite aperiodicity of $\mathcal{S}$, then follows immediately by considering the set $\{x_i : i \in L\} \times K$, where K is constructed as in Remark 3.1. □

Without specifying, $\mathcal{S}$ is always (a, λ)-Ruelle expanding in this paper. We use the notation $x \ll y$, $x \gg y$, $x \asymp y$ to indicate that there exists a positive constant C such that $x \leq Cy$, $x \geq Cy$, $C^{-1}y \leq x \leq Cy$, respectively.

4. *Quotients of Ruelle operators*

In this section, we introduce a family of quotients of Ruelle operators, which will act as strict contractions on the set of probability measures. It provides an effective construction of the relevant measures, whereas a normalization of the Ruelle operators through invariant functions has no dynamical significance in the setting of semi-groups or sequential dynamics due to purely functorial reasons, as noted in Remark 6.6 below.

To begin with, let $\varphi_i : X \to \mathbb{R}$, $i \in \mathcal{W}$ be a continuous function. We also call φ_i a potential. Define for a finite word $v = i_1 i_2 \ldots i_k \in \mathcal{W}^k$,

$$\varphi_v(x) := \varphi_{i_1}(x) + \varphi_{i_2}(T_{i_1}(x)) + \cdots + \varphi_{i_k}(T_{i_1 \ldots i_{k-1}}(x)).$$

Then the Ruelle operator L_v is defined by

$$L_v(f)(x) := \sum_{T_v(y)=x} e^{\varphi_v(y)} f(y)$$

for f in a suitable function space. Note that it follows from $T_v \circ T_u = T_{uv}$ that $L_v \circ L_u = L_{uv}$ for any two finite words u, v. We now define the adequate function space. For $\alpha \in (0, 1]$ and $f : X \to \mathbb{R}$, the Hölder coefficient $D_\alpha(f)$ is

$$D_\alpha(f) := \sup_{x,y \in X, x \neq y} \frac{|f(x) - f(y)|}{d(x, y)^\alpha}$$

and the space of α-Hölder functions $\mathcal{H}_\alpha^*$ is

$$\mathcal{H}_\alpha^* := \{f : D_\alpha(f) < \infty\}.$$

Let $\mathcal{H}_\alpha$ denote the subspace of bounded functions in $\mathcal{H}_\alpha^*$. It is well known that $\mathcal{H}_\alpha$ is a Banach space with respect to the norm $\|\cdot\| := \|\cdot\|_\infty + D_\alpha(\cdot)$. We are now in position to specify the class of potentials considered here.

Definition 4.1. We refer to φ_i as a *α-Hölder potential* if $\varphi_i \in \mathcal{H}_\alpha^*$. Moreover, for any finite word v, we say that φ_v is a *summable potential* if $\|L_v(\mathbf{1})\|_\infty < \infty$.

Suppose φ_i is α-Hölder for every $i \in \mathcal{W}$. We shall estimate distortion of φ_v. Due to the (a, λ)-Ruelle expanding property, for $v = i_1 \ldots i_k \in \mathcal{W}^k$ and $x, y, \tilde{x} \in X$ with $d(x, y) < a$ and $T_v(\tilde{x}) = x$, there exists a unique point $\tilde{y} \in T_v^{-1}(y) \cap B_a^v(\tilde{x})$. Moreover,

$$d(\tilde{x}, \tilde{y}) < \lambda^k d(x, y), \quad d(T_{i_1 \ldots i_j}(\tilde{x}), T_{i_1 \ldots i_j}(\tilde{y})) < \lambda^{k-j} d(x, y), \quad 1 \leq j < k.$$

https://doi.org/10.1017/etds.2022.60 Published online by Cambridge University Press

Hence, the inverse branch

$$(T_v)_{\tilde{x}}^{-1} : B_a(x) \to B_a^v(\tilde{x}), \quad y \mapsto \tilde{y} \tag{4.1}$$

is well defined and contracts the distance at every intermediate step by λ. It follows that for any pair x, y with $d(x, y) < a$, there is a bijection from $T_v^{-1}(x)$ to $T_v^{-1}(y)$ given by

$$\tilde{x} \mapsto \tilde{y}_{\tilde{x}} := (T_v)_{\tilde{x}}^{-1}(y). \tag{4.2}$$

Now Hölder continuity implies that whenever $d(x, y) < a$,

$$|\varphi_v(\tilde{x}) - \varphi_v(\tilde{y}_{\tilde{x}})| \leq \frac{\max_{i \in \mathcal{W}} D_\alpha(\varphi_i)}{1 - \lambda^\alpha} d(x, y)^\alpha =: C_\varphi d(x, y)^\alpha. \tag{4.3}$$

It follows from a simple argument that L_v maps $\mathcal{H}_\alpha$ to $\mathcal{H}_\alpha$ if φ_v is also summable.

As we are interested in operators that leave invariant the constant function $\mathbf{1}$, define for finite words u, v

$$\mathbb{P}_u^v(f) := \frac{L_v(f \cdot L_u(\mathbf{1}))}{L_{uv}(\mathbf{1})} = \frac{L_{uv}(f \circ T_u)}{L_{uv}(\mathbf{1})}.$$

It is clear from the definition that

$$\mathbb{P}_u^v(\mathbf{1}) = \mathbf{1}.$$

The motivation to consider these families of operators stems from the simple observation that for finite words u, v, w,

$$\mathbb{P}_{uv}^w \circ \mathbb{P}_u^v(f) = \frac{L_w(\mathbb{P}_u^v(f) \cdot L_{uv}(\mathbf{1}))}{L_{uvw}(\mathbf{1})} = \frac{L_w(L_v(f \cdot L_u(\mathbf{1})))}{L_{uvw}(\mathbf{1})} = \mathbb{P}_u^{vw}(f).$$

Hence, with

$$\mathbb{P}^w(f) := L_w(f)/L_w(\mathbf{1}),$$

for a sequence of finite words $v_1, \ldots v_k$,

$$\mathbb{P}^{v_1 \ldots v_k} = \mathbb{P}_{v_1 \ldots v_{k-1}}^{v_k} \circ \mathbb{P}_{v_1 \ldots v_{k-2}}^{v_{k-1}} \circ \cdots \circ \mathbb{P}_{v_1 v_2}^{v_3} \circ \mathbb{P}_{v_1}^{v_2} \circ \mathbb{P}^{v_1}. \tag{4.4}$$

As a first result, we obtain $\mathcal{H}_\alpha$-invariance of these quenched operators.

LEMMA 4.1. *$\mathbb{P}_u^v$ is a bounded operator on $\mathcal{H}_\alpha$. Furthermore, for $f \in \mathcal{H}_\alpha$ and x, y with $d(x, y) < a$,*

$$|\mathbb{P}_u^v(f)(x) - \mathbb{P}_u^v(f)(y)| \leq C_\varphi(2\|f\|_\infty + \lambda^{|v|} D_\alpha(f)) d(x, y)^\alpha. \tag{4.5}$$

Proof. Following verbatim the proof of Lemma 2.1 in [3], one obtains that for x, y with $d(x, y) < a$,

$$|L_v(f L_u(\mathbf{1}))(x) - L_v(f L_u(\mathbf{1}))(y)| \leq C_\varphi L_{uv}(\mathbf{1})(x)(\|f\|_\infty + \lambda^{|v|} D_\alpha(f)) d(x, y)^\alpha.$$

The estimate (4.5) follows from this as in [3]. It remains to show that the operators are bounded and leave invariant $\mathcal{H}_\alpha$. As $\mathbb{P}_u^v$ maps positive functions to positive functions and $\mathbb{P}_u^v(\mathbf{1}) = \mathbf{1}$, we have $\|\mathbb{P}_u^v(f)\|_\infty \leq \|f\|_\infty$. Furthermore, by considering the cases

https://doi.org/10.1017/etds.2022.60 Published online by Cambridge University Press

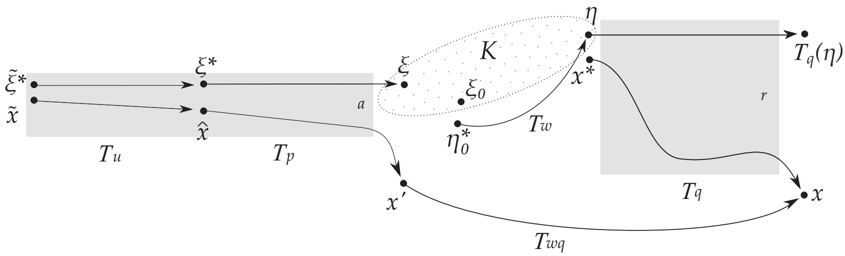

FIGURE 2. Selection of preimages.

$d(x, y) < a$ and $d(x, y) \geq a$ separately, we obtain

$$D_\alpha(\mathbb{P}_u^v(f)) \leq \max\{C_\varphi(2\|f\|_\infty + \lambda^{|v|} D_\alpha(f)), 2a^{-\alpha}\|f\|_\infty\},$$

which proves that $\mathbb{P}_u^v : \mathcal{H}_\alpha \to \mathcal{H}_\alpha$ is a well-defined and bounded operator. □

We observe that Lemma 4.1, which requires Hölder continuity of the potentials and no further assumption on topological irreducibility, is one of the principal ingredients to prove that the duals of the previous operators act as contractions on the space of probabilities. The other ingredient is the following result for which finite aperiodicity is essential.

LEMMA 4.2. *Suppose that $\mathcal{S}$ is jointly topologically mixing and finitely aperiodic, and that every φ_i is α-Hölder and summable. Then $L_v(\mathbf{1})(x) \asymp L_v(\mathbf{1})(y)$, that is, there exists $C > 0$ such that $1/C < L_v(\mathbf{1})(x)/L_v(\mathbf{1})(y) < C$ for all finite words v and $x, y \in X$.*

Proof. First, note that for any $x, y \in X$ with $d(x, y) < a$ and any finite word v, the bijection of equation (4.2) and the estimate (4.3) imply that $L_v(\mathbf{1})(x) \asymp L_v(\mathbf{1})(y)$.

Suppose $\mathcal{S}$ is n-finitely aperiodic. Let K be a finite set and $r > 0$ be given by finite aperiodicity. It follows from the Ruelle expanding property and joint topological mixing that there exists $m \in \mathbb{N}$ such that for all $\xi, \eta \in K$ and $|w| \geq m$, there exists $\eta^* \in X$ with $T_w(\eta^*) = \eta$ and $d(\eta^*, \xi) < a$.

We now show the lemma for any $x, y \in X$ and all finite words v with $|v| > 2n + m$. Take such a finite word v, we will select preimages of x as follows, illustrated in Figure 2.

Decompose $v = upwq$, where u, w, p, q are finite words and $|p| = |q| = n$, $|w| = m$. Note that

$$L_v(\mathbf{1})(x) = L_{wq}(L_{up}(\mathbf{1}))(x) \leq \sup_{i \in \mathcal{W}} \|L_i(\mathbf{1})\|_\infty^{n+m} \sup_{x' \in T_{wq}^{-1}(x)} L_{up}(\mathbf{1})(x').$$

Fix $x' \in T_{wq}^{-1}(x)$. For any $\tilde{x} \in T_{up}^{-1}(x')$, let $\hat{x} = T_u(\tilde{x})$. There exist by condition (1) of finite aperiodicity, $\xi \in K$ and $\xi^* \in T_p^{-1}(\xi)$ such that $d_p(\hat{x}, \xi^*) < a$. Let $\tilde{\xi}^* = (T_u)_{\tilde{x}}^{-1}(\xi^*)$, the inverse branch defined in equation (4.1). Then using equation (4.3),

$$e^{\varphi_{up}(\tilde{x})} = e^{\varphi_u(\tilde{x})} e^{\varphi_p(\hat{x})} \leq e^{C_\varphi a^\alpha + \varphi_u(\tilde{\xi}^*)} e^{na^\alpha + \varphi_p(\xi^*)} = e^{C_\varphi a^\alpha + na^\alpha} e^{\varphi_{up}(\tilde{\xi}^*)}.$$

Because $d_{up}(\tilde{x}, \tilde{\xi}^*) < a$ and $T_{up}(\tilde{\xi}^*) = \xi$, one has $\tilde{x} = (T_{up})^{-1}_{\tilde{\xi}^*}(x')$ and $\tilde{\xi}^* = (T_{up})^{-1}_{\tilde{x}}(\xi)$. Therefore, different $\tilde{x}$ is associated to different $\tilde{\xi}^*$, so that

$$L_{up}(\mathbf{1})(x') = \sum_{\tilde{x} \in T_{up}^{-1}(x')} e^{\varphi_{up}(\tilde{x})} \ll \sum_{\tilde{x} \in T_{up}^{-1}(x')} e^{\varphi_{up}(\tilde{\xi}^*)} \leq \sum_{\xi \in K} L_{up}(\mathbf{1})(\xi).$$

Hence,

$$L_v(\mathbf{1})(x) \ll \sum_{\xi \in K} L_{up}(\mathbf{1})(\xi).$$

However, there exist by condition (2) of finite aperiodicity, a preimage $x^* \in T_q^{-1}(x)$ and $\eta \in K$ such that $d(x^*, \eta) < a$ and $\eta \in B_r^q(x^*)$. As $d(x^*, \eta) < a$, we know that $L_{upw}(\mathbf{1})(x^*) \asymp L_{upw}(\mathbf{1})(\eta)$. Then,

$$L_v(\mathbf{1})(x) \geq e^{\varphi_q(x^*)} L_{upw}(\mathbf{1})(x^*) \gg e^{\varphi_q(\eta) - nr^\alpha} L_{upw}(\mathbf{1})(\eta) \gg L_{upw}(\mathbf{1})(\eta).$$

The last estimate holds because $q \in \mathcal{W}^n$ and $\eta \in K$ both range over finite sets. Now for any $\xi \in K$, one can find $\eta^* \in T_w^{-1}(\eta)$ such that $d(\eta^*, \xi) < a$, then find such a η_0^* for ξ_0 that achieves $\max_{\xi \in K} L_{up}(\mathbf{1})(\xi)$. Then, $L_{up}(\mathbf{1})(\xi_0) \asymp L_{up}(\mathbf{1})(\eta_0^*)$ and

$$\begin{aligned} L_{upw}(\mathbf{1})(\eta) &= \sum_{\eta^* \in T_w^{-1}(\eta)} e^{\varphi_w(\eta^*)} L_{up}(\mathbf{1})(\eta^*) \geq e^{\varphi_w(\eta_0^*)} L_{up}(\mathbf{1})(\eta_0^*) \\ &\gg e^{\varphi_w(\eta_0^*)} L_{up}(\mathbf{1})(\xi_0) \gg L_{up}(\mathbf{1})(\xi_0). \end{aligned}$$

The last estimate holds because φ_w is continuous, $\eta_0^* \in \overline{B_a(\xi_0)}$, $\xi_0 \in K$ and $w \in \mathcal{W}^m$ range over finite sets. Therefore,

$$L_v(\mathbf{1})(x) \gg \max_{\xi \in K} L_{up}(\mathbf{1})(\xi).$$

All the constants absorbed into $\ll$ or $\gg$ are determined by $\mathcal{S}, \varphi, K, m, n$ (essentially by $\mathcal{S}$ and φ), in particular independent of v, x, y. It follows from the above estimates that $L_v(\mathbf{1})(x) \asymp L_v(\mathbf{1})(y)$ for any $x, y \in X$.

Lastly, when $|v| \leq 2n + m$, take any finite word $|v'| > 2n + m$, then for any $x \in X$,

$$\begin{aligned} L_{v'v}(\mathbf{1})(x) = L_v(L_{v'}(\mathbf{1}))(x) &= \sum_{\tilde{x} \in T_v^{-1}(x)} e^{\varphi(\tilde{x})} L_{v'}(\mathbf{1})(\tilde{x}) \asymp \sum_{\tilde{x} \in T_v^{-1}(x)} e^{\varphi(\tilde{x})} L_{v'}(\mathbf{1})(x) \\ &= L_v(\mathbf{1})(x) L_{v'}(\mathbf{1})(x) \end{aligned}$$

by the already-proven case. So $L_v(\mathbf{1})(x) \asymp L_{v'v}(\mathbf{1})(x)/L_{v'}(\mathbf{1})(x)$, and hence for any $x, y \in X$, $L_v(\mathbf{1})(x) \asymp L_v(\mathbf{1})(y)$. □

5. *Contraction in the Wasserstein distance*

Let $\mathcal{M}_1(X)$ refer to the space of Borel probability measures on *X*. Recall that the Wasserstein distance W of $\mu, \nu \in \mathcal{M}_1(X)$ defined by

$$W(\mu, \nu) := \inf \left\{ \int d(x, y)\, dP : P \in \Pi(\mu, \nu) \right\}$$

https://doi.org/10.1017/etds.2022.60 Published online by Cambridge University Press

is a compatible metric with weak convergence, where $\Pi(\mu, \nu)$ refers to the couplings of μ and ν, that is, the set of probability measures on $X \times X$ with marginal distributions μ and ν. Moreover, by Kantorovich's duality,

$$W(\mu, \nu) = \sup \left\{ \left| \int f \, d(\mu - \nu) \right| : \sup_{x \neq y} \frac{|f(x) - f(y)|}{d(x, y)} \leq 1 \right\}.$$

Let $\mathbb{P}_u^{v*}$ denote the dual operator of $\mathbb{P}_u^v$ on $\mathcal{M}_1(X)$. To obtain a contraction of $W(\mathbb{P}_u^{v*}(\cdot), \mathbb{P}_u^{v*}(\cdot))$, the estimates of Lemma 4.1 indicate that for a-close measures, one should consider $(d(x, y))^\alpha$ instead of $d(x, y)$. However, for distant measures, the method of proof below based on an idea in [17] (see also [3, 23, 31, 32]) requires a truncated distance. We consider

$$d^*(x, y) := \min\{1, \Delta \, d(x, y)^\alpha\}, \quad \Delta := \max\{4C_\varphi, a^{-\alpha}\}. \tag{5.1}$$

Observe that, by construction, $d(x, y) < a$ whenever $d^*(x, y) < 1$. To see that d^* is a metric, observe that the triangle inequality follows from $x^\alpha + y^\alpha \geq (x + y)^\alpha$ for $x, y \geq 0$ and $0 < \alpha \leq 1$, which is an inequality that easily can be deduced from the concavity of $x \mapsto x^\alpha$. The remaining assertion that $d^*(x, y) = 0$ if and only if $x = y$ is trivial.

We now introduce the space of d^*-Lipschitz functions. To do so, recall that the Lipschitz coefficient is defined by $D_{d^*}(f) := \sup\{|f(x) - f(y)|/d^*(x, y) : x \neq y\}$ and that f is a bounded Lipschitz continuous function with respect to d^* if and only if $\|f\| := \|f\|_\infty + D_{d^*}(f) < \infty$. To identify these functions in terms of the metric d, set

$$\overline{D}(f) := \max\{ \sup_{x,y \in X} |f(x) - f(y)|, D_\alpha^{loc}(f)/\Delta\},$$

where

$$D_\alpha^{loc}(f) := \sup \left\{ \frac{|f(x) - f(y)|}{d(x, y)^\alpha} : x, y \in X, 0 < d(x, y) < \Delta^{-1/\alpha} \right\}.$$

Now observe that it follows from the construction that $\overline{D}(f) = D_{d^*}(f)$, $\overline{D}(f) \leq 2\|f\|_\infty + \Delta^{-1} D_\alpha(f)$ and $D_\alpha(f) \leq \Delta \overline{D}(f)$. Hence, the norms $\| \cdot \|_\infty + D_\alpha^{loc}(\cdot)$ and $\| \cdot \|_\infty + D_{d^*}(\cdot)$ are equivalent. In particular, by Kantorovich's duality, the Wasserstein metric $\overline{W}$ with respect to d^* is characterized through local Hölder continuous functions with respect to d by

$$\overline{W}(\mu, \nu) = \sup \left\{ \left| \int f \, d(\mu - \nu) \right| : \overline{D}(f) \leq 1 \right\}.$$

THEOREM 5.1. *Suppose that $\mathcal{S}$ is jointly topologically mixing and a finitely aperiodic Ruelle expanding semigroup, and that every potential φ_i is α-Hölder and summable. Then there exist $k_0 \in \mathbb{N}$ and $s \in (0, 1)$ such that for all finite words u, v with $|v| \geq k_0$ and $\nu_1, \nu_2 \in \mathcal{M}_1(X)$ and f with $\overline{D}(f) < \infty$,*

$$\overline{W}(\mathbb{P}_u^{v*}(\nu_1), \mathbb{P}_u^{v*}(\nu_2)) \leq s^n \overline{W}(\nu_1, \nu_2),$$
$$\overline{D}(\mathbb{P}_u^v(f)) \leq s^n \overline{D}(f).$$

Remark 5.2. Under the additional hypothesis that X is compact, the condition of finite aperiodicity is automatically satisfied.

https://doi.org/10.1017/etds.2022.60 Published online by Cambridge University Press

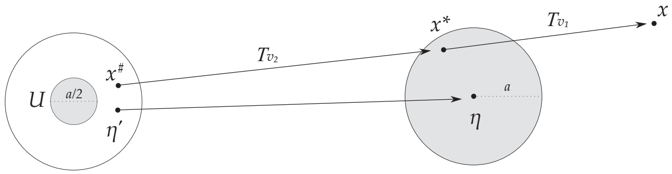

FIGURE 3. The map $x \mapsto x^\#$.

Proof. As in [17], we first prove the assertions for Dirac measures and then extend the partial result by optimal transport to arbitrary probability measures.

(1) *Local contraction*. Assume that $d^*(x, y) < 1$ and that f is d^*-Lipschitz continuous. Since $d(x, y) < a$ as soon as $d^*(x, y) < 1$, Lemma 4.1 gives that

$$|\mathbb{P}_u^v(f)(x) - \mathbb{P}_u^v(f)(y)| \leq (2C_\varphi \|f\|_\infty + \lambda^{|v|} D_\alpha^{loc}(f))(d(x, y))^\alpha.$$

Furthermore, as $\mathbb{P}_u^v(\mathbf{1}) = \mathbf{1}$, one may suppose without loss of generality that $\inf f = 0$, and therefore, $\|f\|_\infty \leq \overline{D}(f)$. Dividing by Δ and choosing k_0 such that $\lambda^{k_0} \leq 1/4$, it follows that for v with $|v| \geq k_0$,

$$|\mathbb{P}_u^v(f)(x) - \mathbb{P}_u^v(f)(y)| \leq \left(\frac{\|f\|_\infty}{2} + \frac{D_\alpha^{loc}(f)}{4\Delta}\right) d^*(x, y) \leq \frac{3\overline{D}(f)}{4} d^*(x, y).$$

Hence, by Kantorovich's duality,

$$\overline{W}(\mathbb{P}_u^{v*}(\delta_x), \mathbb{P}_u^{v*}(\delta_y)) \leq \tfrac{3}{4} d^*(x, y) = \tfrac{3}{4} \overline{W}(\delta_x, \delta_y).$$

(2) *Global contraction*. If $d^*(x, y) = 1$, an upper bound for $\overline{W}$ can be obtained by construction of a coupling based on finite aperiodicity. To do so, fix an open set U of diameter smaller than $a/2$. Suppose $\mathcal{S}$ is n_1-finitely aperiodic and K, r are given by finite aperiodicity. As $\mathcal{S}$ is jointly topologically mixing, one can find n_2 such that $T_w(U) \cap B_a(\xi) \neq \emptyset$ for all $w \in \mathcal{W}^{n_2}$ and $\xi \in K$ and that $\lambda^{n_2} < 1/8$. Choose n_3 large such that $C_{n_3} := \Delta(a\lambda^{n_3})^\alpha < 1/2$. Let $k_0 = n_1 + n_2 + n_3$.

Let $n \geq k_0$. For $v \in \mathcal{W}^n$, write $v = v_3 v_2 v_1$, where $|v_1| = n_1$, $|v_2| = n_2$ and $|v_3| \geq n_3$. For any $x \in X$, we will select a preimage $x^\#$ in $T_{v_2 v_1}^{-1}(x)$ as below, illustrated in Figure 3.

Let $\eta \in K$ and $x^* \in X$ be given by condition (2) of finite aperiodicity so that $T_{v_1}(x^*) = x$, $d(x^*, \eta) < a$ and $x^* \in B_r^{v_1}(\eta)$. Now the choice of n_2 and Ruelle expanding property allow us to find a preimage $\eta' \in T_{v_2}^{-1}(\eta)$ such that $\eta' \in B_{a/8}(U)$. Use the Ruelle expanding property again to find a preimage $x^\# \in T_{v_2}^{-1}(x^*)$ such that $x^\# \in B_{a/8}(\eta') \subset B_{a/4}(U)$. One has $|\varphi_{v_2}(x^\#) - \varphi_{v_2}(\eta')| \leq C_\varphi a^\alpha$ by equation (4.3), so that

$$|\varphi_{v_2 v_1}(x^\#) - \varphi_{v_2 v_1}(\eta')| \leq C_\varphi a^\alpha + n_1 r^\alpha \max_{i \in \mathcal{W}} D_\alpha(\varphi_i),$$

and hence

$$e^{\varphi_{v_2 v_1}(x^\#)} \asymp e^{\varphi_{v_2 v_1}(\eta')} = e^{\varphi_{v_2}(\eta')} e^{\varphi_{v_1}(\eta)}.$$

https://doi.org/10.1017/etds.2022.60 Published online by Cambridge University Press

Since η' lies in a fixed bounded region $B_{a/8}(U)$ and φ is continuous and $\eta \in K$, $v_1 \in \mathcal{W}^{n_1}$, $v_2 \in \mathcal{W}^{n_2}$ range over finite sets, one concludes that for all $x \in X$, $v_1 \in \mathcal{W}^{n_1}$, $v_2 \in \mathcal{W}^{n_2}$,

$$e^{\varphi_{v_2 v_1}(x^{\#})} \asymp 1. \tag{5.2}$$

For any pair $(x, y) \in X^2$, find as before $x^{\#}, y^{\#} \in B_{a/4}(U)$. Then $d(x^{\#}, y^{\#}) < a$. As stated in equation (4.2), there is a bijection $\tilde{x} \mapsto \tilde{y}$ from $T_{v_3}^{-1}(x^{\#})$ to $T_{v_3}^{-1}(y^{\#})$. Pair $(\tilde{x}, \tilde{y})$ together by this bijection and set a subprobability measure on X^2,

$$Q_{(x,y)} := \min\left\{ \sum_{(\tilde{x},\tilde{y})} \frac{e^{\varphi_v(\tilde{x})} L_u(\mathbf{1})(\tilde{x})}{L_{uv}(\mathbf{1})(x)}\, \delta_{(\tilde{x},\tilde{y})},\ \sum_{(\tilde{x},\tilde{y})} \frac{e^{\varphi_v(\tilde{y})} L_u(\mathbf{1})(\tilde{y})}{L_{uv}(\mathbf{1})(y)}\, \delta_{(\tilde{x},\tilde{y})} \right\}.$$

Note that $Q_{(x,y)}(X^2) = Q_{(x,y)}(\{(z_1, z_2) : d(z_1, z_2) < a\lambda^{|v_3|}\})$. For any $A \subset X$,

$$Q_{(x,y)}(A \times X) \le \sum_{T_v(z)=x} \frac{e^{\varphi_v(z)} \mathbf{1}_A \cdot L_u(\mathbf{1})(z)}{L_{uv}(\mathbf{1})(x)} = \frac{L_v(\mathbf{1}_A \cdot L_u(\mathbf{1}))}{L_{uv}(\mathbf{1})}(x) = \mathbb{P}_u^{v*}(\delta_x)(A)$$

and similarly $Q_{(x,y)}(X \times A) \le \mathbb{P}_u^{v*}(\delta_y)(A)$. Hence, there exists a further subprobability measure R such that $P := Q_{(x,y)} + R \in \Pi(\mathbb{P}_u^{v*}(\delta_x), \mathbb{P}_u^{v*}(\delta_y))$ (see, e.g. [17]). Therefore, due to the choice of n_3,

$$\begin{aligned}
\overline{W}(\mathbb{P}_u^{v*}(\delta_x), \mathbb{P}_u^{v*}(\delta_y)) &\le \int d^*(z_1, z_2)\, dP \\
&\le \Delta(a\lambda^{|v_3|})^{\alpha} P(\{d(z_1, z_2) < a\lambda^{|v_3|}\}) + P(\{d(z_1, z_2) \ge a\lambda^{|v_3|}\}) \\
&\le 1 - C_{n_3} P(\{d(z_1, z_2) < a\lambda^{|v_3|}\}) \le 1 - C_{n_3} Q_{(x,y)}(X^2).
\end{aligned}$$

To get a lower bound for $Q_{(x,y)}(X^2)$, use equation (5.2) to see

$$\begin{aligned}
Q_{(x,y)}(X^2) &\asymp \min\left\{ \sum_{T_{v_3}(\tilde{x})=x^{\#}} \frac{e^{\varphi_{v_3}(\tilde{x})} L_u(\mathbf{1})(\tilde{x})}{L_{uv}(\mathbf{1})(x)},\ \sum_{T_{v_3}(\tilde{y})=y^{\#}} \frac{e^{\varphi_{v_3}(\tilde{y})} L_u(\mathbf{1})(\tilde{y})}{L_{uv}(\mathbf{1})(y)} \right\} \\
&= \min\left\{ \frac{L_{uv_3}(\mathbf{1})(x^{\#})}{L_{uv}(\mathbf{1})(x)},\ \frac{L_{uv_3}(\mathbf{1})(y^{\#})}{L_{uv}(\mathbf{1})(y)} \right\}.
\end{aligned}$$

Applying Lemma 4.2, we get that for any $\xi_0 \in K$,

$$Q_{(x,y)}(X^2) \asymp \frac{1}{L_{v_2 v_1}(\mathbf{1})(\xi_0)} \ge \min\{(L_w(\mathbf{1})(\xi))^{-1} : \xi \in K, w \in \mathcal{W}^{n_1+n_2}\} > 0.$$

Hence, there is a lower bound $N \le Q_{(x,y)}(X^2)$, independent of $x, y \in X$ and $v \in \mathcal{W}^n$. Therefore, increasing n_3 so that $C_{n_3} N < 1$ if needed,

$$\overline{W}(\mathbb{P}_u^{v*}(\delta_x), \mathbb{P}_u^{v*}(\delta_y)) \le 1 - C_{n_3} N = (1 - C_{n_3} N) d^*(x, y) = (1 - C_{n_3} N)\overline{W}(\delta_x, \delta_y).$$

Combining part (1) with part (2) of the proof and letting $t := \max\{3/4, 1 - C_{n_3} N\} < 1$, we obtain that there exists k_0 such that for all finite words u, v with $|v| \ge k_0$ and $x, y \in X$,

$$\overline{W}(\mathbb{P}_u^{v*}(\delta_x), \mathbb{P}_u^{v*}(\delta_y)) \le t\,\overline{W}(\delta_x, \delta_y).$$

https://doi.org/10.1017/etds.2022.60 Published online by Cambridge University Press

Using Kantorovich's duality, for f with $\overline{D}(f) \leq 1$, it follows that

$$|\mathbb{P}_u^v(f)(x) - \mathbb{P}_u^v(f)(y)| = \left| \int f d\mathbb{P}_u^{v*}(\delta_x) - \int f \, d\mathbb{P}_u^{v*}(\delta_y) \right| \leq t.$$

(3) *Contraction for arbitrary probability measures*. The extension to arbitrary probability measures is a standard application of optimal transport and omitted as the proof is a straightforward adaption of [17], [31] or [23]. We obtain that for any finite words u, v with $|v| \geq k_0$ and any probability measures ν_1, ν_2,

$$\overline{W}(\mathbb{P}_u^{v*}(\nu_1), \mathbb{P}_u^{v*}(\nu_2)) \leq t\overline{W}(\nu_1, \nu_2).$$

(4) *Iteration.* By the iteration rules given in equation (4.4), the theorem follows for $s = t^{1/2k_0}$. □

6. *Conformal measures, quenched exponential decay and continuity*

From now on, we always assume that $\mathcal{S}$ is jointly topologically mixing and finitely aperiodic and every potential φ_i is α-Hölder and summable, so that Theorem 5.1 holds. It has immediate consequences for the existence and regularity of two types of compact sets of probability measures, which are canonical generalizations of conformal measures and equilibrium states to the context of semigroups.

6.1. *One-sided dynamics.* Denote by $\Sigma = \{i_1 i_2 \ldots : i_1, i_2, \ldots \in \mathcal{W}\}$ the set of infinite words and by $\theta(i_1 i_2 \ldots) = i_2 i_3 \ldots$ the shift map. For an infinite word $\omega = i_1 i_2 \ldots \in \Sigma$ and $k \in \mathbb{N}$, let

$$[\omega]_k := i_1 \ldots i_k \in \mathcal{W}^k.$$

The first family of measures is constructed as follows, which generalizes the notion of conformal measures.

PROPOSITION 6.1. *For any finite word u, infinite word ω and measure $\nu \in \mathcal{M}_1(X)$, the limit*

$$\mu_{u,\omega} := \lim_{l \to \infty} \mathbb{P}_u^{[\omega]_l *}(\nu)$$

exists and is independent of ν. Furthermore, with k_0 and s given by Theorem 5.1, the following statements hold.

(1) *For $k \geq k_0$ and any $\omega, \tilde{\omega} \in \Sigma$ with $[\omega]_k = [\tilde{\omega}]_k$, $\overline{W}(\mu_{u,\omega}, \mu_{u,\tilde{\omega}}) \leq s^k$.*

(2) *For $k \geq k_0$ and $f \in \mathcal{H}_\alpha$,*

$$\left\| \mathbb{P}_u^{[\omega]_k}(f) - \int f \, d\mu_{u,\omega} \right\| \leq 2s^k \overline{D}(f).$$

(3) *Let $\mu_\omega := \mu_{\emptyset,\omega}$, then*

$$\mu_{u\omega} = \mathbb{P}^{u*}(\mu_{u,\omega}), \quad \mu_{u,\omega} = \mu_{u\omega} \circ T_u^{-1}.$$

If v is a finite word,

$$\mu_{u,v\omega} = \mathbb{P}_u^{v*}(\mu_{uv,\omega}).$$

https://doi.org/10.1017/etds.2022.60 Published online by Cambridge University Press

(4) *Let* $\lambda_{u,\omega} := \int L_u(\mathbf{1})\, d\mu_\omega$, *then*

$$L_u^*(\mu_\omega) = \lambda_{u,\omega}\mu_{u\omega},$$

and if v *is a finite word,*

$$\lambda_{uv,\omega} = \lambda_{u,v\omega}\lambda_{v,\omega}.$$

(5) *The measures* $\mu_{u,\omega}$ *and* μ_ω *are absolutely continuous to each other and*

$$h_{u,\omega} := \frac{d\mu_{u,\omega}}{d\mu_\omega} = \lambda_{u,\omega}^{-1} L_u(\mathbf{1}).$$

Proof. For probability measures ν, $\tilde{\nu}$ on X and $l > k \geq k_0$, Theorem 5.1 implies

$$\overline{W}(\mathbb{P}_u^{[\omega]_k *}(\nu), \mathbb{P}_u^{[\omega]_l *}(\tilde{\nu})) = \overline{W}(\mathbb{P}_u^{[\omega]_k *}(\nu), \mathbb{P}_u^{[\omega]_k *} \circ \mathbb{P}_{u[\omega]_k}^{[\theta^k\omega]_{l-k} *}(\tilde{\nu})) \leq s^k.$$

Hence, $\{\mathbb{P}_u^{[\omega]_k *}(\nu)\}_{k\geq k_0}$ is a Cauchy sequence and $\mu_{u,\omega} := \lim_k \mathbb{P}_u^{[\omega]_k *}(\nu)$ exists and is independent of ν. This, in particular, implies the estimate in item (1). To show item (2), it suffices to consider $\nu = \delta_x$. If $k \geq k_0$, we have that

$$\left| \mathbb{P}_u^{[\omega]_k}(f)(x) - \int f\, d\mu_{u,\omega} \right| \leq \overline{D}(f) s^k.$$

The estimate in item (2) then follows from this combined with Theorem 5.1.

The second part of item (3) follows from

$$\int \mathbb{P}_u^v(f)\, d\mu_{uv,\omega} = \lim_{k\to\infty} \mathbb{P}_{uv}^{[\omega]_k} \circ \mathbb{P}_u^v(f)(x) = \lim_{k\to\infty} \mathbb{P}_u^{v[\omega]_k}(f)(x) = \int f\, d\mu_{u,v\omega}.$$

The first part of item (3) follows from this and

$$\begin{aligned}\int f\, d\mu_{u,\omega} &= \lim_{k\to\infty} \frac{L_{[\omega]_k}(f L_u(\mathbf{1}))(x)}{L_{u[\omega]_k}(\mathbf{1})(x)} = \lim_{k\to\infty} \frac{L_{u[\omega]_k}(f \circ T_u)(x)}{L_{u[\omega]_k}(\mathbf{1})(x)} \\ &= \int f \circ T_u\, d\mu_{u\omega} = \int f\, d\mu_{u\omega} \circ T_u^{-1}.\end{aligned}$$

Item (4) holds because

$$\begin{aligned}\int L_u(f)\, d\mu_\omega &= \lim_{k\to\infty} \frac{L_{[\omega]_k}(L_u(f))(x)}{L_{[\omega]_k}(\mathbf{1})(x)} = \lim_{k\to\infty} \frac{L_{u[\omega]_k}(f)(x)}{L_{u[\omega]_k}(\mathbf{1})(x)} \cdot \frac{L_{u[\omega]_k}(\mathbf{1})(x)}{L_{[\omega]_k}(\mathbf{1})(x)} \\ &= \int f\, d\mu_{u\omega} \int L_u(\mathbf{1})\, d\mu_\omega\end{aligned}$$

and

$$\lambda_{uv,\omega}\mu_{uv\omega} = L_{uv}^*(\mu_\omega) = L_u^* L_v^*(\mu_\omega) = L_u^*(\lambda_{v,\omega}\mu_{v\omega}) = \lambda_{v,\omega}\lambda_{u,v\omega}\mu_{uv\omega}.$$

Item (5) follows from

$$\int f\, d\mu_{u,\omega} = \lim_{k\to\infty} \frac{L_{[\omega]_k}(\mathbf{1})(x)}{L_{u[\omega]_k}(\mathbf{1})(x)} \cdot \frac{L_{[\omega]_k}(f L_u(\mathbf{1}))(x)}{L_{[\omega]_k}(\mathbf{1})(x)} = \frac{1}{\lambda_{u,\omega}} \int f L_u(\mathbf{1})\, d\mu_\omega. \qquad \square$$

https://doi.org/10.1017/etds.2022.60 Published online by Cambridge University Press

Remark 6.2. Recall that a probability measure ν is (T_w, φ_w)-*conformal*, where w is a finite word, if there exists $c > 0$ such that $L_w^*(\nu) = c\nu$. Consider $\overline{w} := ww \ldots \in \Sigma$ and $\mu_{\overline{w}} = \mu_{\emptyset,\overline{w}}$ given by Proposition 6.1. By item (4) of the same proposition, $L_w^*(\mu_{\overline{w}}) = \lambda_{w,\overline{w}}\mu_{\overline{w}}$, hence $\mu_{\overline{w}}$ is conformal. Moreover, item (1) and $\mu_{u\overline{w}} \circ T_u^{-1} = \mu_{u,\overline{w}}$ imply

$$\{\mu_{u,\omega} : \omega \in \Sigma\} = \overline{\left\{\mu_{u\overline{w}} \circ T_u^{-1} : w \in \bigcup_{k\geq 1} \mathcal{W}^k\right\}}.$$

As Σ is compact and $\omega \mapsto \mu_{u,\omega}$ is Lipschitz continuous by statement (1) of Proposition 6.1, $\{\mu_{u,\omega} : \omega \in \Sigma\}$ is compact. It is also worth mentioning that item (1) ensures that for any finite word u, the family $\Sigma \ni \omega \mapsto \mu_{u,\omega}$ is Hölder continuous. Finally, the fact that any two asymptotic limits are equivalent (recall item (5)) will be useful to provide an application to characterize the boundary of a semigroup action in §9.

6.2. *Two-sided compositions.* We shall find a second family of probabilities which generalizes the notions of invariant measures and equilibrium states. To attain that goal, despite the fact that the underlying dynamics is not invertible, we need to consider forward iterations of maps determined by two-sided sequences. Let Σ^- refer to the set of left-infinite words, that is, $\Sigma^- = \{\ldots i_2 i_1 : i_1, i_2, \ldots \in \mathcal{W}\}$, and for $k \in \mathbb{N}$ and $\sigma = \ldots i_2 i_1 \in \Sigma^-$, define

$${}_k[\sigma] := i_k \ldots i_2 i_1 \in \mathcal{W}^k.$$

PROPOSITION 6.3. *For any $\sigma \in \Sigma^-$, $\omega \in \Sigma$ and $\nu \in \mathcal{M}_1(X)$, the limit*

$$\mu_{\sigma,\omega} := \lim_{k,l\to\infty} {\mathbb{P}_{{}_k[\sigma]}^{[\omega]_l}}^*(\nu)$$

exists and is independent of ν. Furthermore, with k_0 and s given by Theorem 5.1, the following statements hold.

(1) *For k, l with $k \wedge l \geq k_0$ and $\sigma, \tilde{\sigma} \in \Sigma^-, \omega, \tilde{\omega} \in \Sigma$ with ${}_k[\sigma] = {}_k[\tilde{\sigma}]$, $[\omega]_l = [\tilde{\omega}]_l$, $\overline{W}(\mu_{\sigma,\omega}, \mu_{\tilde{\sigma},\tilde{\omega}}) \leq s^{k\wedge l}$.*

(2) *For k, l with $k \wedge l \geq k_0$ and $f \in \mathcal{H}_\alpha$,*

$$\left\| \mathbb{P}_{{}_k[\sigma]}^{[\omega]_l}(f) - \int f\, d\mu_{\sigma,\omega} \right\| \leq 2s^{k\wedge l}\overline{D}(f).$$

(3) *For a finite word u, $\mu_{\sigma u,\omega} = \mu_{\sigma,u\omega} \circ T_u^{-1}$.*

(4) *The measures $\mu_{\sigma,\omega}$ and μ_ω are absolutely continuous to each other and $h_{\sigma,\omega} := d\mu_{\sigma,\omega}/d\mu_\omega$ satisfies*

$$\|h_{{}_k[\sigma],\omega} - h_{\sigma,\omega}\| \ll s^k,$$

where μ_ω and $h_{{}_k[\sigma],\omega}$ are as given in the previous proposition.

Proof. As a consequence of Proposition 6.1(2), Lemmas 4.1 and 4.2, for any finite word u, infinite word $\omega \in \Sigma$ and $l \geq k_0$, we have that

$$\|L_{u[\omega]_l}(\mathbf{1})/L_{[\omega]_l}(\mathbf{1}) - \lambda_{u,\omega}\| \leq s^l\overline{D}(L_u(\mathbf{1})) \leq Cs^l\lambda_{u,\omega}, \tag{6.1}$$

https://doi.org/10.1017/etds.2022.60 Published online by Cambridge University Press

for some $C > 0$. Hence, for finite words $v \in \mathcal{W}^k$, $w \in \mathcal{W}^l$, $k \geq k_0$ and f Hölder continuous,

$$\begin{aligned}
&|\mathbb{P}_v^w(f) - \mathbb{P}_{uv}^w(f)| \\
&\quad \leq \left| \frac{L_w(fL_v(\mathbf{1}))}{L_{vw}(\mathbf{1})} - \frac{L_w(fL_{uv}(\mathbf{1}))}{\lambda_{u,\overline{vw}} L_{vw}(\mathbf{1})} \right| + \left| \frac{L_w(fL_{uv}(\mathbf{1}))}{\lambda_{u,\overline{vw}} L_{vw}(\mathbf{1})} - \frac{L_w(fL_{uv}(\mathbf{1}))}{L_{uvw}(\mathbf{1})} \right| \\
&\quad \leq \frac{L_w(|f|L_v(\mathbf{1})|1 - L_{uv}(\mathbf{1})/\lambda_{u,\overline{vw}} L_v(\mathbf{1})|)}{L_{vw}(\mathbf{1})} + \frac{L_w(|f|L_{uv}(\mathbf{1}))}{L_{uvw}(\mathbf{1})} \left| \frac{L_{uvw}(\mathbf{1})}{\lambda_{u,\overline{vw}} L_{vw}(\mathbf{1})} - 1 \right| \\
&\quad \leq C(\mathbb{P}_v^w(|f|)s^k + \mathbb{P}_{uv}^w(|f|)s^{k+l}),
\end{aligned}$$

where we used the notation $\overline{u} := (uu\ldots)$ to denote the periodic word formed by u blocks. Now assume that ν and $\tilde{\nu}$ are probability measures and f is Hölder continuous with $\overline{D}(f) \leq 1$ and $\inf_{x \in X} f(x) = 0$. In particular, $\|f\|_\infty \leq 1$. By the above and Proposition 6.1, for $\sigma, \tilde{\sigma} \in \Sigma^-$ and $\omega, \tilde{\omega} \in \Sigma$ such that ${}_k[\sigma] = {}_k[\tilde{\sigma}]$, $[\omega]_l = [\tilde{\omega}]_l$ and $k \wedge l \geq k_0$,

$$\begin{aligned}
&\left| \int \mathbb{P}_{k[\sigma]}^{[\omega]_l}(f)\, d\nu - \int \mathbb{P}_{k[\tilde{\sigma}]}^{[\tilde{\omega}]_l}(f)\, d\tilde{\nu} \right| \\
&\quad \leq \int \left| \mathbb{P}_{k[\sigma]}^{[\omega]_l}(f) - \mathbb{P}_{k[\tilde{\sigma}]}^{[\omega]_l}(f) \right| d\nu + \left| \int \mathbb{P}_{k[\tilde{\sigma}]}^{[\omega]_l}(f)\, d\nu - \int \mathbb{P}_{k[\tilde{\sigma}]}^{[\omega]_l}(f)\, d\tilde{\nu} \right| \\
&\quad \leq C(2\|\mathbb{P}_{k[\sigma]}^{[\omega]_l}(f)\|_\infty s^k + \|\mathbb{P}_{k[\sigma]}^{[\omega]_l}(f)\|_\infty s^{k+l} + \|\mathbb{P}_{k[\tilde{\sigma}]}^{[\omega]_l}(f)\|_\infty s^{k+l}) + 2s^l \\
&\quad \leq 2C(s^k + s^{k+l}) + 2s^l \ll s^{k \wedge l}.
\end{aligned}$$

Hence, by Kantorovich's duality and completeness of the space of probability measures, $\lim_{k,l \to \infty} \mathbb{P}_{k[\sigma]}^{[\omega]_l\,*}(\nu)$ exists, is independent of ν and the estimate in part (1) holds. Part (2) is an immediate consequence of part (1), and the proof of part (3) follows as in Proposition 6.1. Proposition 6.1(5) indicates that $h_{\sigma,\omega}$ is the limit of $h_{k[\sigma],\omega}$ and by the first argument in Proposition 2.2 in [3], it follows that $\|h_{k[\sigma],\omega} - h_{l[\sigma],\omega}\|_\infty \ll s^{k \wedge l}$. Then the argument in there can be easily adapted to obtain exponential convergence with respect to $\|\cdot\|_{d^*}$ in part (4). □

Remark 6.4. The first part of the above proposition implies that the map $(\sigma, \omega) \mapsto \mu_{\sigma,\omega}$ is Lipschitz continuous with respect to the metric

$$d((\sigma, \omega), (\tilde{\sigma}, \tilde{\omega})) := \min\{s^{k \wedge l} : {}_k[\sigma] = {}_k[\tilde{\sigma}], [\omega]_l = [\tilde{\omega}]_l\}.$$

In particular, the image of each compact subset of $\Sigma^- \times \Sigma$ is a compact subset of the space of probability measures.

Moreover, by fixing an order on $\mathcal{W}$, the associated adic flow h_t on $\Sigma^- \times \Sigma$ is uniquely ergodic (see [15]) and, in particular, for any Hölder continuous $f : X \to \mathbb{R}$, the continuity of $(\sigma, \omega) \to \int f\, d\mu_{\sigma,\omega}$ implies that

$$\frac{1}{T} \int_0^T \int f(x)\, d\mu_{h_t(\sigma,\omega)}(x)\, dt \xrightarrow{T \to \infty} \iint f(x)\, d\nu_{\sigma,\omega}(x)\, dm(\sigma, \omega)$$

uniformly, where m refers to the Parry measure (or measure of maximal entropy). The analogue of this statement holds for $\omega \to \int f\, d\mu_{\omega,\omega}$ and Birkhoff sums with respect to

https://doi.org/10.1017/etds.2022.60 Published online by Cambridge University Press

the odometer on Σ, or with respect to uniformly ergodic adic flows or adic transformations acting on compact subsets of $\Sigma^- \times \Sigma$ or Σ, respectively.

The result provides the following link to invariant measures and equilibrium states. A finite word w generates a periodic infinite word $\overline{w} := (ww\ldots) \in \Sigma$ and a periodic left-infinite word $\underline{w} := (\ldots ww) \in \Sigma^-$. Then, by Proposition 6.3, the measure $\mu_{\underline{w},\overline{w}}$ is T_w-invariant, $d\mu_{\underline{w},\overline{w}} = h_{\underline{w},\overline{w}}\, d\mu_{\overline{w}}$ and

$$L_w(h_{\underline{w},\overline{w}}) = \lambda_{w,\overline{w}} h_{\underline{w},\overline{w}}.$$

Here, $\lambda_{w,\overline{w}}$ is given as in Proposition 6.1.

The following result identifies $\mu_{\underline{w},\overline{w}}$ as the unique equilibrium state of T_w with respect to the Hölder potential φ_w. Note that the statement avoids the notion of pressure as X might be non-compact. However, if X is compact, then $\log \lambda_{w,\overline{w}}$ is equal to the pressure [28] and one obtains the usual notion of equilibrium state. In the proposition, $H_\mu(T_w)$ refers to Kolmogorov's entropy.

PROPOSITION 6.5.

$$\begin{aligned}\log \lambda_{w,\overline{w}} =& H_{\mu_{\underline{w},\overline{w}}}(T_w) + \int \varphi_w \, d\mu_{\underline{w},\overline{w}} \\ &= \sup \left\{ H_\nu(T_w) + \int \varphi_w \, d\nu : \nu \in \mathcal{M}_1(X),\, \nu = \nu \circ T_w^{-1} \right\}.\end{aligned}$$

Furthermore, $\mu_{\underline{w},\overline{w}}$ is the unique measure which realizes the supremum.

Proof. As T_w is Ruelle expanding, the restriction $T_w|_U$ to a ball U of radius a is bimeasurable. Hence, $A \mapsto \mu_{\underline{w},\overline{w}} \circ T_w(A)$ defines a measure on U which is, as a consequence of Propositions 6.1 and 6.3, absolutely continuous with respect to $\mu_{\underline{w},\overline{w}}|_U$. Hence, $J_{\mu_{\underline{w},\overline{w}}} := d\mu_{\underline{w},\overline{w}} \circ T_w / d\mu_{\underline{w},\overline{w}}$ is a well-defined function on X, sometimes referred to as the Jacobian of T_w with respect to $\mu_{\underline{w},\overline{w}}$. In fact, it follows from the construction of $\mu_{\underline{w},\overline{w}}$ that $J_{\mu_{\underline{w},\overline{w}}} = \exp(-\tilde{\varphi}_w)$, where

$$\tilde{\varphi}_w := \varphi_w + \log h_{\underline{w},\overline{w}} - \log h_{\underline{w},\overline{w}} \circ T_w - \log \lambda_{w,\overline{w}}.$$

By construction, $J_{\mu_{\underline{w},\overline{w}}} = \exp(-\tilde{\varphi}_w)$ and, as T_w is Ruelle expanding, Rokhlin's formula for entropy (see, e.g. Theorem 9.7.3 in [35]) implies that

$$\begin{aligned}H_{\mu_{\underline{w},\overline{w}}}(T_w) &= \int \log J_{\mu_{\underline{w},\overline{w}}} \, d\mu_{\underline{w},\overline{w}} \\ &= \log \lambda_{w,\overline{w}} - \int (\varphi_w + \log h_{\underline{w},\overline{w}} - \log h_{\underline{w},\overline{w}} \circ T_w) \, d\mu_{\underline{w},\overline{w}} \\ &= \log \lambda_{w,\overline{w}} - \int \varphi_w \, d\mu_{\underline{w},\overline{w}}.\end{aligned}$$

This proves the first identity. Now suppose that ν is an invariant probability measure with $H_\nu(T_w) + \int \varphi_w \, d\nu \geq \log \lambda_{w,\overline{w}}$. Then, by Rokhlin's formula, the invariance of ν and the

https://doi.org/10.1017/etds.2022.60 Published online by Cambridge University Press

definition of the transfer operator of T_w with respect to ν, denoting by $J_\nu = d\nu \circ T_w / d\nu$,

$$\begin{aligned} 0 &\leq H_\nu(T_w) + \int \varphi_w \, d\nu - \log \lambda_{w,\overline{w}} \\ &= \int (\log J_\nu + \varphi_w + \log h_{\underline{w},\overline{w}} - \log h_{\underline{w},\overline{w}} \circ T_w - \log \lambda_{w,\overline{w}}) \, d\nu \\ &= \int \log \frac{J_\nu}{J_{\mu_{\underline{w},\overline{w}}}} \, d\nu = \int \sum_{T_w(y)=x} \frac{1}{J_\nu(y)} \log \frac{J_\nu(y)}{J_{\mu_{\underline{w},\overline{w}}}(y)} \, d\nu(x). \end{aligned}$$

As ν is invariant, it follows that $\sum_{T_w(y)=x} 1/J_\nu(y) = 1$ for all $x \in X$. Hence, by Jensens's inequality,

$$0 \leq H_\nu(T_w) + \int \varphi_w \, d\nu - \log \lambda_{w,\overline{w}} \stackrel{*}{\leq} \int \log \sum_{T_w(y)=x} \frac{1}{J_\nu(y)} \frac{J_\nu(y)}{J_{\mu_{\underline{w},\overline{w}}}(y)} \, d\nu(x) = 0.$$

Moreover, equality holds in $(*)$ if and only if $J_\nu(y)/J_{\mu_{\underline{w},\overline{w}}}(y) = 1$ almost surely. □

Remark 6.6. By usual normalization procedure, replacing the potential φ_w with $\tilde{\varphi}_w$, one then obtains a new operator $\tilde{L}_w$ with $\tilde{L}_w(\mathbf{1}) = \mathbf{1}$, that is, $\tilde{L}_w$ is normalized and $\tilde{L}_w^*(\mu_{\underline{w},\overline{w}}) = \mu_{\underline{w},\overline{w}}$. In particular, part (2) of Proposition 6.1 applied to the semigroup generated by T_w implies that $\tilde{L}_w$ has a spectral gap. However, the construction depends on the specific periodic word $\overline{w}$ and is in general not functorial, that is, $\tilde{L}_{vw} \neq \tilde{L}_w \circ \tilde{L}_v$.

7. *Annealed exponential decay*

So far, we have considered only quenched operators, which are determined by iterations in $\mathcal{S}$ tracked by certain finite words and their limiting behaviour. As stated in the introduction, another objective is to study annealed operators, which are averages of all the quenched operators tracked by finite words of given lengths. To be more precise, suppose that the one-sided full shift of finite alphabet (Σ, θ) is endowed with a non-singular probability measure ρ. For every $k \in \mathbb{N}$, define the *averaged transfer operator*

$$\mathcal{A}_k(f)(x) := \int_\Sigma L_{[\omega]_k}(f)(x) \, d\rho(\omega)$$

for $f \in \mathcal{H}_\alpha$. One can do so for more general shifts, but we keep Σ to be a topological mixing subshift of finite type for simplicity. Naturally, one would need some properties of the shift space (Σ, θ, ρ) to study the operator $\mathcal{A}_k$. We summarize them below.

Since ρ is non-singular, for a finite word u, let $p_u : \Sigma \to \mathbb{R}_+$ be defined by

$$p_u(\omega) := \frac{d\rho}{d\rho \circ \theta^{|u|}}(u\omega), \quad \omega \in \Sigma.$$

With the usual distance given on the shift, denote by $\mathcal{H}(\Sigma)$ the space of Hölder continuous functions on Σ and by $\mathcal{C}(\Sigma)$ the space of continuous functions on Σ. Recall that $\lambda_{u,\omega} = \int L_u(\mathbf{1}) \, d\mu_\omega$, as in Proposition 6.1. Note that $\log \lambda_{i,\cdot} \in \mathcal{H}(\Sigma)$ by Proposition 6.1. Suppose that $\log p_i \in \mathcal{H}(\Sigma)$ as well. Define a linear operator ι acting on $\mathcal{C}(\Sigma)$ by

$$\iota(g)(\omega) := \sum_{i \in \mathcal{W}} \lambda_{i,\omega} p_i(\omega) g(i\omega), \quad g \in \mathcal{C}(\Sigma).$$

https://doi.org/10.1017/etds.2022.60 Published online by Cambridge University Press

As $u \mapsto p_u$ and $u \mapsto \lambda_{u,\omega}$ are multiplicative cocycles with respect to θ, it can be shown that for every $k \in \mathbb{N}$,

$$\iota^k(g)(\omega) = \sum_{u \in \mathcal{W}^k} \lambda_{u,\omega} p_u(\omega) g(u\omega).$$

In view of the duality with θ, we have that for any $g_1, g_2 \in \mathcal{C}(\Sigma)$,

$$\int \iota^k(g_1) \cdot g_2 \, d\rho = \int \lambda_{[\omega]_k, \theta^k \omega} \cdot g_1 \cdot g_2 \circ \theta^k \, d\rho. \tag{7.1}$$

Since $\log \lambda_{i,\omega}$ and $\log p_i$ are both Hölder continuous, Ruelle's Perron–Frobenius theorem implies that there are $\beta > 0$, $m \in \mathcal{M}_1(\Sigma)$ and $g_o \in \mathcal{C}(\Sigma)$, $g_o > 0$ such that

$$\iota^* m = \beta m, \quad \iota(g_o) = \beta g_o, \quad m(g_o) = 1. \tag{7.2}$$

Furthermore, there exists $t \in (0, 1)$ such that for any $g \in \mathcal{H}(\Sigma)$ and $k \in \mathbb{N}$,

$$\left\| \beta^{-k} \iota^k(g) - g_o \int g \, dm \right\|_\Sigma \ll t^k \|g\|_\Sigma, \tag{7.3}$$

where $\|\cdot\|_\Sigma = D_\Sigma(\cdot) + \|\cdot\|_\infty$, the sum of the Hölder norm and the supremum norm over the shift. Note that g_o is uniformly bounded from above and away from 0 as Σ is compact.

Remark 7.1. If $(i, \omega) \mapsto \lambda_{i,\omega}$ is constant, then $m = \rho$. Moreover, if ρ is invariant, then $g_o = 1$. If ρ is a Bernoulli measure, then $\mathcal{A}_k = (\mathcal{A}_1)^k$ for every $k \geq 1$. In this case, annealed transfer operators were studied in [2]. Note that $\mathcal{A}_l \circ \mathcal{A}_k = \mathcal{A}_{l+k}$ if and only if ρ is Bernoulli. Averaged transfer operators were also considered in [6] in the special case that ρ is a Bernoulli measure and all potentials φ_i are equal.

Remark 7.2. The associated skew product

$$F : X \times \Sigma \to X \times \Sigma, \quad (x, i_1 i_2 \ldots) \mapsto (T_{i_1}(x), i_2 i_3 \ldots)$$

reflects the time evolution along a given path in Σ with a distribution on the space of possible paths, that is, the probability of the event of applying $T \in \mathcal{S}$ in time n is given by $\rho(\{\omega \in \Sigma : F^n(\cdot, \omega) = (T(\cdot), \theta^n(\omega))\})$.

We proceed to prove that the family $\{\mathcal{A}_n\}$ has exponential decay of correlations. Fix $k_0 \in \mathbb{N}$ and $s \in (0, 1)$, as given in Theorem 5.1. With m defined as in equation (7.2), let $\pi \in \mathcal{M}_1(X)$ be given by

$$d\pi := d\mu_\omega dm(\omega).$$

For $f \in \mathcal{H}_\alpha$, let

$$\|f\|_m := \|\mu_\cdot(|f|)\|_\infty$$

be the supremum norm with respect to m of the map $\omega \mapsto \mu_\omega(|f|)$ over the shift.

THEOREM 7.3. *Suppose the Ruelle expanding semigroup $\mathcal{S}$ is jointly topologically mixing and finitely aperiodic, and that every potential φ_i is α-Hölder and summable. Suppose that*

https://doi.org/10.1017/etds.2022.60 Published online by Cambridge University Press

every $\log p_i$, $i \in \mathcal{W}$ *is Hölder continuous on* Σ. *Then there exists* $r \in (0, 1)$ *such that for all* $f \in \mathcal{H}_\alpha$ *and* $n \geq 2k_0$,

$$\left| \frac{\mathcal{A}_n(f)(x)}{\mathcal{A}_n(\mathbf{1})(x)} - \int f \, d\pi \right| \ll r^n(\overline{D}(f) + \|f\|_m).$$

Moreover, there exists a positive function $h \in \mathcal{H}_\alpha$ *such that for all* $f \in \mathcal{H}_\alpha$ *and* $n \geq 2k_0$,

$$\left| \frac{\mathcal{A}_n(f)(x)}{\beta^n h(x)} - \int f \, d\pi \right| \ll r^n(\overline{D}(f) + \|f\|_m),$$

with $\beta > 0$ *given by equation (7.2).*

Proof. In the first step of the proof, we derive the first decay. Proposition 6.1 implies that for any $n \geq 2k_0$, $\omega \in \Sigma$ and $x \in X$, $f \in \mathcal{H}_\alpha$,

$$|L_{[\omega]_n}(f)(x) - \mu_\omega(f) L_{[\omega]_n}(\mathbf{1})(x)| \ll s^n \overline{D}(f) L_{[\omega]_n}(\mathbf{1})(x).$$

After integration, it yields that

$$\left| \mathcal{A}_n(f)(x) - \int \mu_\omega(f) L_{[\omega]_n}(\mathbf{1})(x) \, d\rho(\omega) \right| \ll s^n \overline{D}(f) \mathcal{A}_n(\mathbf{1})(x). \tag{7.4}$$

It remains to analyse $\int \mu_\omega(f) L_{[\omega]_n}(\mathbf{1}) \, d\rho(\omega)$ as $n \to \infty$. To do so, write $n = k + l$ with $l = [n/2] + 1$. Observe that by equation (6.1),

$$|L_{[\omega]_n}(\mathbf{1}) - \lambda_{[\omega]_k, \theta^k\omega} L_{[\theta^k\omega]_l}(\mathbf{1})| \ll s^l \lambda_{[\omega]_k, \theta^k\omega} L_{[\theta^k\omega]_l}(\mathbf{1}). \tag{7.5}$$

Note that it follows from Proposition 6.1 that $\omega \mapsto \mu_\omega(f)$ is Hölder continuous on Σ and its Hölder coefficient is bounded by a constant times $\overline{D}(f)$. Hence,

$$\begin{aligned}
&\left| \int \mu_\omega(f) L_{[\omega]_n}(\mathbf{1}) \, d\rho(\omega) - \int \mu_\omega(f) \lambda_{[\omega]_k, \theta^k\omega} L_{[\theta^k\omega]_l}(\mathbf{1}) \, d\rho(\omega) \right| \\
&\quad \ll s^l \int \mu_\omega(|f|) \lambda_{[\omega]_k, \theta^k\omega} L_{[\theta^k\omega]_l}(\mathbf{1}) \, d\rho(\omega) \\
&\quad \overset{\text{equation (7.1)}}{=} s^l \int \iota^k(\mu_\omega(|f|)) \cdot L_{[\omega]_l}(\mathbf{1}) \, d\rho(\omega) \\
&\quad = s^l \int (\beta^{-k} g_o^{-1} \iota^k(\mu_\omega(|f|)) - \pi(|f|) + \pi(|f|)) \cdot \iota^k(g_o) L_{[\omega]_l}(\mathbf{1}) \, d\rho(\omega) \\
&\quad \overset{\text{equation (7.3)}}{\ll} s^l(t^k(\overline{D}(f) + \|f\|_m) + \pi(|f|)) \int \iota^k(g_o) L_{[\omega]_l}(\mathbf{1}) \, d\rho(\omega) \\
&\quad \overset{\text{equation (7.1)}}{=} s^l(t^k(\overline{D}(f) + \|f\|_m) + \pi(|f|)) \int g_o \cdot \lambda_{[\omega]_k, \theta^k\omega} L_{[\theta^k\omega]_l}(\mathbf{1}) \, d\rho(\omega) \\
&\quad \overset{\text{equation (7.5)}}{\ll} s^l(t^k(\overline{D}(f) + \|f\|_m) + \pi(|f|)) \int L_{[\omega]_n}(\mathbf{1}) \cdot g_o \, d\rho(\omega) \\
&\quad \ll s^l(t^k \overline{D}(f) + \|f\|_m) \mathcal{A}_n(\mathbf{1}).
\end{aligned}$$

Observe that in the previous estimate, we have also shown that

$$\int \iota^k(g_o) L_{[\omega]_l}(\mathbf{1}) \, d\rho(\omega) \ll \mathcal{A}_n(\mathbf{1}). \tag{7.6}$$

https://doi.org/10.1017/etds.2022.60 Published online by Cambridge University Press

Then one can extract $\pi(f)$ by

$$\begin{aligned}
&\left| \int \mu_\omega(f)\lambda_{[\omega]_k,\theta^k\omega} L_{[\theta^k\omega]_l}(\mathbf{1})\, d\rho(\omega) - \pi(f)\int \lambda_{[\omega]_k,\theta^k\omega} L_{[\theta^k\omega]_l}(\mathbf{1})\, d\rho(\omega)\right| \\
&\overset{\text{equation (7.1)}}{=} \left| \int \iota^k(\mu_\omega(f)) L_{[\omega]_l}(\mathbf{1})\, d\rho(\omega) - \pi(f)\int \iota^k(1) L_{[\omega]_l}(\mathbf{1})\, d\rho(\omega)\right| \\
&= \left| \int ((\beta^{-k} g_o^{-1}\iota^k(\mu_\omega(f)) - \pi(f)) - (\beta^{-k} g_o^{-1}\iota^k(1) - 1)\pi(f))\iota^k(g_o) L_{[\omega]_l}(\mathbf{1})\, d\rho(\omega)\right| \\
&\overset{\text{equation (7.3)}}{\ll} t^k(\overline{D}(f) + \|f\|_m)\int \iota^k(g_o) L_{[\omega]_l}(\mathbf{1})\, d\rho(\omega) \ll t^k(\overline{D}(f) + \|f\|_m)\mathcal{A}_n(\mathbf{1}).
\end{aligned}$$

Finally, equation (7.5) induces that

$$\left| \pi(f)\int \lambda_{[\omega]_k,\theta^k\omega} L_{[\theta^k\omega]_l}(\mathbf{1})\, d\rho(\omega) - \pi(f)\mathcal{A}_n(\mathbf{1})\right| \ll s^l|\pi(f)|\mathcal{A}_n(\mathbf{1}).$$

Combining the above estimates, one obtains that

$$\left| \int \mu_\omega(f) L_{[\omega]_n}(\mathbf{1})\, d\rho(\omega) - \pi(f)\mathcal{A}_n(\mathbf{1})\right| \ll (t^k\overline{D}(f) + t^k\|f\|_m + s^l\|f\|_m)\mathcal{A}_n(\mathbf{1}).$$

The first statement now follows from equation (7.4) with $r = \max\{\sqrt{s}, \sqrt[3]{t}\}$.

We now proceed with proving the existence of h. To do so, let

$$\tilde{\mathcal{A}}_n(x) := \int L_{[\omega]_n}(\mathbf{1})(x)\cdot g_o(\omega)\, d\rho(\omega).$$

We first show that $\tilde{I}_n(x) := \beta^{-n}\tilde{\mathcal{A}}_n(x)$ converges uniformly and exponentially fast to a positive function $h(x) \in \mathcal{H}_\alpha$.

It follows from equation (7.5) that for any $n = k + l$ with $l \geq k_0$,

$$L_{[\omega]_n}(\mathbf{1}) \asymp \lambda_{[\omega]_k,\theta^k\omega} L_{[\theta^k\omega]_l}(\mathbf{1}),$$

so that

$$\tilde{\mathcal{A}}_n \asymp \int \lambda_{[\omega]_k,\theta^k\omega} L_{[\theta^k\omega]_l}(\mathbf{1})\cdot g_o\, d\rho \overset{\text{equation (7.1)}}{=} \int \iota^k(g_o) L_{[\omega]_l}(\mathbf{1})\, d\rho = \beta^k\tilde{\mathcal{A}}_l,$$

and hence, $\tilde{I}_n \asymp \tilde{I}_l$, especially $\tilde{I}_n \asymp \tilde{I}_{k_0}$ for all $n \geq k_0$. Since equation (7.5) also implies that

$$|\tilde{\mathcal{A}}_n - \beta^k\tilde{\mathcal{A}}_l| \ll s^l\beta^k\tilde{\mathcal{A}}_l,$$

one has

$$|\tilde{I}_n - \tilde{I}_l| \ll s^l\tilde{I}_l.$$

Hence, $\{\tilde{I}_n(\cdot)\}$ is a Cauchy sequence. Denote the limit of $\tilde{I}_n(x)$ by $h(x)$. Then $\tilde{I}_n(x)$ converges uniformly to $h(x)$ since for $n \geq l \geq k_0$,

$$|\tilde{I}_n - \tilde{I}_l| \ll s^l\tilde{I}_{k_0} \ll s^l.$$

Then because $\tilde{I}_n$ are all Hölder, h is Hölder as well. That h is positive and $\|h\|_\infty$ is finite can be seen from $h \asymp \tilde{I}_{k_0}$. To see that the rate of convergence is exponential, for $n \geq k_0$, choose $j \in \mathbb{N}$ such that $|\tilde{I}_{jn} - h| \le s^n$, then

$$|\tilde{I}_n - h| \le |\tilde{I}_n - \tilde{I}_{2n}| + \cdots + |\tilde{I}_{(j-1)n} - \tilde{I}_{jn}| + |\tilde{I}_{jn} - h| \ll s^n.$$

Moreover, Lemma 4.2 infers that $\inf_{x \in X} \tilde{I}_{k_0}(x) > 0$, and so are $\tilde{I}_n$ for $n \ge k_0$ and so is h. It follows that $\tilde{I}_n / h$ converges to 1 uniformly and exponentially fast.

Next we show that $I_n(x) := \beta^{-n} \mathcal{A}_n(\mathbf{1})(x)$ also tends to $h(x)$. For $n = k + l$ with $l \ge k_0$, because

$$\left| \mathcal{A}_n(\mathbf{1}) - \int \iota^k(1) L_{[\omega]_l}(\mathbf{1}) \, d\rho \right| \ll s^l \int \iota^k(1) L_{[\omega]_l}(\mathbf{1}) \, d\rho$$

obtained from integrating equation (7.5) and because

$$\begin{aligned} \left| \int (\iota^k(1) - \iota^k(g_o)) L_{[\omega]_l}(\mathbf{1}) \, d\rho \right| &= \left| \int (\beta^{-k} g_o^{-1} \iota^k(1) - 1) \iota^k(g_o) L_{[\omega]_l} \, d\rho \right| \\ &\overset{\text{equation (7.3)}}{\le} t^k \int \iota^k(g_o) L_{[\omega]_l} \, d\rho = t^k \beta^k \tilde{A}_l, \end{aligned}$$

one can deduce that

$$|\mathcal{A}_n(\mathbf{1}) - \beta^k \tilde{\mathcal{A}}_l| \ll (s^l + t^k) \beta^k \tilde{A}_l,$$

and hence

$$|I_n - \tilde{I}_l| \ll (s^l + t^k) \tilde{I}_l,$$

so that

$$|I_n - h| \ll (s^l + t^k) h.$$

Lastly, applying Theorem 7.3, one has that for all $f \in \mathcal{H}_\alpha$ and $n \ge 2k_0$,

$$\begin{aligned} |\beta^{-n} \mathcal{A}_n(f) - \pi(f) h| &\le \beta^{-n} |\mathcal{A}_n(f) - \pi(f) \mathcal{A}_n(\mathbf{1})| + \pi(f) |\beta^{-n} \mathcal{A}_n(\mathbf{1}) - h| \\ &\ll r^n (\overline{D}(f) + \|f\|_m) I_n + \pi(f) |I_n - h| \\ &\ll r^n (\overline{D}(f) + \|f\|_m) h. \end{aligned}$$

The second assertion on the decay follows from this. □

The next result reveals an annealed version of the decay of correlations.

THEOREM 7.4. *Now suppose that the assumptions of the above theorem hold and that, in addition, ρ is θ-invariant. Then there exist a probability measure $\tilde{\pi}$ on $\Sigma \times X$, $r \in (0, 1)$ and $k_1 \in \mathbb{N}$ such that*

$$\begin{aligned} &\left| \int \sum_{v \in \mathcal{W}^n} \mathbf{1}_{[v]}(\omega) f(T_v(x)) g(x) \, d\mu_\omega(x) \, d\rho(\omega) - \int f \, d\tilde{\pi} \int g \, d\mu_\omega \, d\rho \right| \\ &\quad \le r^n \int |f| \, d\mu_\omega \, d\rho \left(\overline{D}(g) + \int |g| \, d\mu_\omega \, d\rho \right) \end{aligned}$$

for all $n \ge k_1$, $g \in \mathcal{H}_\alpha$ and $f : X \to \mathbb{R}$ integrable with respect to $d\mu_\omega(x)\, d\rho(\omega)$.

https://doi.org/10.1017/etds.2022.60 Published online by Cambridge University Press

Proof. For $\omega = (\omega_1 \omega_2 \ldots) \in \Sigma$, set $\lambda_{n,\omega} := \lambda_{\omega_1 \ldots \omega_n, \theta^n \omega}$ and $h_{n,\omega} := h_{\omega_1 \ldots \omega_n, \theta^n \omega}$, where $\lambda_{\cdot}$ and $h_{\cdot}$ are given by Proposition 6.1. Moreover, Proposition 6.1 and Lemma 4.2 imply for n sufficiently large that

$$\begin{aligned}
\int \sum_{v \in \mathcal{W}^n} \mathbf{1}_{[v]} f \circ T_v g \, d\mu_\omega \, d\rho &= \int \sum_{v \in \mathcal{W}^n} \mathbf{1}_{[v]} f \frac{L_v(g)}{\lambda_{n,\omega}} \, d\mu_{\theta^n \omega} \, d\rho \\
&= \int \sum_{v} \mathbf{1}_{[v]} f \mu_\omega(g) \frac{L_v(\mathbf{1})}{\lambda_{n,\omega}} \, d\mu_{\theta^n \omega} \, d\rho \pm 2 s^n \overline{D}(g) \int \sum_{v} \mathbf{1}_{[v]} |f| \frac{L_v(\mathbf{1})}{\lambda_{n,\omega}} \, d\mu_{\theta^n \omega} \, d\rho \\
&= \int \sum_{v \in \mathcal{W}^n} \mathbf{1}_{[v]} f \mu_\omega(g) h_{n,\omega} \, d\mu_{\theta^n \omega} \, d\rho \pm C s^n \overline{D}(g) \int \sum_{v \in \mathcal{W}^n} \mathbf{1}_{[v]} |f| \, d\mu_{\theta^n \omega} \, d\rho \\
&= \int f \mu_\omega(g) h_{n,\omega} \, d\mu_{\theta^n \omega} \, d\rho \pm C s^n \overline{D}(g) \int |f| \, d\mu_\omega \, d\rho,
\end{aligned} \tag{7.7}$$

where $C/2$ is given by Lemma 4.2, and the last equality follows from θ-invariance of ρ. Now assume that n is even and $n = 2m$. Then, by item (4) of Proposition 6.3, there exists C such that

$$\begin{aligned}
&\int f \mu_\omega(g) h_{n,\omega} \, d\mu_{\theta^n \omega} \, d\rho \\
&\quad = \int f \mu_\omega(g) h_{m,\theta^m \omega} \, d\mu_{\theta^n \omega} \, d\rho \pm C s^m \int \mu_{\theta^n \omega}(|f|) |\mu_\omega(g)| \, d\rho.
\end{aligned}$$

However, as $\omega \to \mu_\omega(g)$ is Lipschitz continuous by Proposition 6.1, the exponential decay of correlations, say with rate $t \in (0, 1)$ and the same constant $C > 0$, applied to the error term implies that

$$\begin{aligned}
&\int f \mu_\omega(g) h_{m,\theta^m \omega} \, d\mu_{\theta^n \omega} \, d\rho \pm C s^m \int \mu_{\theta^n \omega}(|f|) |\mu_\omega(g)| \, d\rho \\
&\quad = \int f \mu_\omega(g) h_{m,\theta^m \omega} \, d\mu_{\theta^n \omega} \, d\rho \pm C^2 s^m \int \mu_\omega(|f|) \, d\rho \int \mu_\omega(|g|) \, d\rho.
\end{aligned} \tag{7.8}$$

A further application of invariance and the exponential decay of correlations of θ to the main term and Lemma 4.2 gives that

$$\begin{aligned}
&\int f \mu_\omega(g) h_{m,\omega} \, d\mu_{\theta^n \omega} \, d\rho = \int \mu_\omega(g) \mu_{\theta^{2m} \omega}(f \, h_{m,\theta^m \omega}) \, d\rho \\
&\quad = \int \mu_\omega(g) \, d\rho \int f h_{m,\omega} \, d\mu_{\theta^m \omega} \, d\rho \pm C^2 t^m \int \mu_\omega(|f|) \, d\rho \overline{D}(g)
\end{aligned} \tag{7.9}$$

Hence, it remains to analyse $\int f h_{m,\omega} \, d\mu_{\theta^m \omega}$. To do so, let $(\hat{\Sigma}, \hat{\theta}, \hat{\rho})$ refer to natural extension of θ. Then, again by item (4) of Proposition 6.3, it follows that

$$\begin{aligned}
&\int f h_{m,\omega} \, d\mu_{\theta^m \omega} \, d\rho(\omega) = \int f h_{m,\omega} \, d\mu_{\theta^m \omega} \, d\hat{\rho}(\tilde{\omega}, \omega) \\
&\quad = \int f h_{\tilde{\omega}_{-m} \cdots \tilde{\omega}_{-1}, \omega} \, d\mu_\omega d\hat{\rho}(\tilde{\omega}, \omega) = \int f h_{\tilde{\omega}, \omega} \, d\mu_\omega \, d\hat{\rho}(\tilde{\omega}, \omega) \pm C s^m \int \mu_\omega(|f|) \, d\rho \\
&\quad = \int f \, d\mu_{\tilde{\omega}, \omega} \, d\hat{\rho}(\tilde{\omega}, \omega) \pm C s^m \int \mu_\omega(|f|) \, d\rho.
\end{aligned} \tag{7.10}$$

Let $d\tilde{\pi}(x) := d\mu_{\tilde{\omega},\omega}(x)d\hat{\rho}(\tilde{\omega}, \omega)$. The theorem now follows by combining equations (7.7), (7.8), (7.9) and (7.10). □

Remark 7.5. As a corollary of the proof, we also obtain an explicit representation of $\tilde{\pi}$. That is, $d\tilde{\pi}(x) := d\mu_{\tilde{\omega},\omega}(x)d\hat{\rho}(\tilde{\omega}, \omega)$, where $\hat{\rho}$ is the natural extension of ρ (which is assumed invariant). In particular, $d\tilde{\pi}$ and $d\mu_\omega\, d\rho(\omega)$ are equivalent measures, even though $d\tilde{\pi}/d\mu_\omega\, d\rho(\omega)$ might be a function depending on ω. However, it is not clear if $\tilde{\pi}$ and π coincide. Furthermore, this representation reveals that in our sequential setting, the measure arising in the annealed version of the decay of correlations is an integral of the pathwise equilibrium measures, as known for the special case where ρ is a Bernoulli measure.

8. *An almost sure invariance principle*

Exponential decay of correlations has many implications on the statistical behaviour of the dynamical system. A large deviation principle, a relativized central limit theorem and laws of iterated logarithm for random dynamical systems generated by expanding dynamics follow from the works by Kifer [21, 22]. For sequential dynamical systems of expanding maps of the interval, first versions of central limit theorems were obtained by Heinrich [19] and Conze and Raugi [9]. We now show an almost sure invariance principle in the setting of Ruelle expanding maps. It is worth mentioning that almost sure invariance principles have been obtained in the context of quenched random dynamical systems (see e.g. [13] and references therein). Let $\mathcal{B}$ be the Borel σ-algebra on X. With respect to the measure $\mu_{uv\omega}$, where u, v are finite words and ω is an infinite word, $\mathbb{P}_u^v$ can be seen as a conditional expectation in the following way.

LEMMA 8.1. *For any $f \in \mathcal{H}_\alpha$,*

$$\mathbb{E}_{\mu_{uv\omega}}(f \circ T_u | T_{uv}^{-1}\mathcal{B}) = \mathbb{P}_u^v(f) \circ T_{uv}.$$

Proof. For any $A \in \mathcal{B}$, using item (3) of Proposition 6.1,

$$\begin{aligned}\int_{T_{uv}^{-1}A} f \circ T_u\, d\mu_{uv\omega} &= \int \mathbf{1}_A \circ T_v \cdot f\, d\mu_{uv\omega} \circ T_u^{-1} = \int \mathbf{1}_A \circ T_v \cdot f\, d\mu_{u,v\omega}\\ &= \int \mathbf{1}_A \circ T_v \cdot f\, d\mathbb{P}_u^{v*}(\mu_{uv,\omega}) = \int \mathbb{P}_u^v(\mathbf{1}_A \circ T_v \cdot f)\, d\mu_{uv,\omega}\\ &= \int \mathbf{1}_A \cdot \mathbb{P}_u^v(f)\, d\mu_{uv,\omega} = \int_A \mathbb{P}_u^v(f)\, d\mu_{uv\omega} \circ T_{uv}^{-1}\\ &= \int_{T_{uv}^{-1}A} \mathbb{P}_u^v(f) \circ T_{uv}\, d\mu_{uv\omega}.\end{aligned}$$

□

The almost sure invariance principle we are going to show is similar to the one in [32] for non-stationary shift. Both are based on the almost sure invariance principle for reverse martingale differences by Cuny and Merlevède.

THEOREM 8.2. [10, Theorem 2.3] *Let $(U_n)_{n\in\mathbb{N}}$ be a sequence of square integrable reverse martingale differences with respect to a non-increasing filtration $(\mathcal{G}_n)_{n\in\mathbb{N}}$. Assume that*

https://doi.org/10.1017/etds.2022.60 Published online by Cambridge University Press

$\sigma_n^2 := \sum_{k=1}^n \mathbb{E}(U_k^2) \to \infty$ *and that* $\sup_n \mathbb{E}(U_n^2) < \infty$. *Assume that*

$$\sum_{k=1}^n (\mathbb{E}(U_k^2|\mathcal{G}_{k+1}) - \mathbb{E}(U_k^2)) = o(\sigma_n^2) \quad \textit{almost surely},$$

$$\sum_{n\geq 1} \sigma_n^{-2t} \mathbb{E}(|U_n|^{2t}) < \infty \quad \textit{for some } 1 \leq t \leq 2.$$

Then, enlarging our probability space if necessary, it is possible to find a sequence $(Z_k)_{k\geq 1}$ *of independent centred Gaussian variables with* $\mathbb{E}(Z_k^2) = \mathbb{E}(U_k^2)$ *such that*

$$\sup_{1\leq k\leq n} \left| \sum_{i=1}^k U_i - \sum_{i=1}^k Z_i \right| = o(\sqrt{\sigma_n^2 \log\log \sigma_n^2}) \quad \textit{almost surely}.$$

We need to make another assumption.

Definition 8.1. An (a, λ)-Ruelle expanding map T is *finitely expanding* if

$$\sup_{\substack{x,y\in X \\ 0<d(x,y)<a}} \frac{d(T(x), T(y))}{d(x, y)} < \infty.$$

We refer to $\mathcal{S}$ as finitely Ruelle expanding if every T_i, $i \in \mathcal{W}$ satisfies this property.

THEOREM 8.3. *Suppose the finitely Ruelle expanding semigroup* $\mathcal{S}$ *is jointly topologically mixing and finitely aperiodic, and that every potential* φ_i *is* α*-Hölder and summable. Suppose* $\omega \in \Sigma$, $f \in \mathcal{H}_\alpha$. *Let* $f_n = f - \int f \circ T_{[\omega]_n}\, d\mu_\omega$ *for every* $n \in \mathbb{N}_0$ *and let* $s_n^2 = \mathbb{E}_{\mu_\omega}(\sum_{k=0}^{n-1} f_k \circ T_{[\omega]_k})^2$ *for* $n \geq 1$. *Assume that*

$$\sum_n s_n^{-4} < \infty. \tag{8.1}$$

Then, enlarging our probability space if necessary, there exists a sequence (Z_n) *of independent centred Gaussian random variables such that*

$$\sup_n \left| \sqrt{\textstyle\sum_{k=0}^{n-1} \mathbb{E}_{\mu_\omega} Z_k^2} - s_n \right| < \infty,$$

$$\sup_{0\leq k\leq n-1} \left| \sum_{i=0}^k f_i \circ T_{[\omega]_i} - \sum_{i=0}^k Z_i \right| = o\left(\sqrt{s_n^2 \log\log s_n^2} \right) \quad \mu_\omega\textit{-almost surely}.$$

Proof. Denote $\mathcal{B}_n = T_{[\omega]_n}^{-1} \mathcal{B}$ for $n \in \mathbb{N}$ and let $\mathcal{B}_0 = \mathcal{B}$, then $\mathcal{B}_n$ is a non-increasing filtration. Let $h_0 = 0$ and define $h_n \in \mathcal{H}_\alpha$ recursively by $h_{n+1} = \mathbb{P}_{[\omega]_n}^{[\theta^n\omega]_1}(f_n + h_n)$. Then equation (4.4) implies that $h_n = \sum_{k=0}^{n-1} \mathbb{P}_{[\omega]_k}^{[\theta^k\omega]_{n-k}} f_k \in \mathcal{H}_\alpha$. It follows from Proposition 6.1 that $\mu_\omega \circ T_{[\omega]_k}^{-1} = \mu_{[\omega]_k, \theta^k\omega}$, then

$$\mathbb{P}_{[\omega]_k}^{[\theta^k\omega]_{n-k}} f_k = \mathbb{P}_{[\omega]_k}^{[\theta^k\omega]_{n-k}} f - \int f \circ T_{[\omega]_k}\, d\mu_\omega = \mathbb{P}_{[\omega]_k}^{[\theta^k\omega]_{n-k}} f - \int f\, d\mu_{[\omega]_k, \theta^k\omega}$$

https://doi.org/10.1017/etds.2022.60 Published online by Cambridge University Press

and that, with $k_0 \in \mathbb{N}$ and $s \in (0, 1)$ given by Theorem 5.1,

$$\|h_n\| \leq \sum_{k=0}^{n-k_0} 2s^{n-k}\overline{D}(f) + \sum_{k=n-k_0+1}^{n-1} \|\mathbb{P}_{[\omega]_k}^{[\theta^k\omega]_{n-k}} f_k\|$$
$$\leq \sum_{k=0}^{n-k_0} 2s^{n-k}\overline{D}(f) + \sum_{k=n-k_0+1}^{n-1} C\|f\| \ll \|f\|,$$

where C is a uniform bound for all $\|\mathbb{P}_u^v\|$ (Lemma 4.1).

Let

$$U_n := f_n \circ T_{[\omega]_n} + h_n \circ T_{[\omega]_n} - h_{n+1} \circ T_{[\omega]_{n+1}}.$$

Here, U_n is $\mathcal{B}_n$-measurable and square integrable. Moreover, apply Lemma 8.1 to get that

$$\mathbb{E}_{\mu_\omega}(U_n|\mathcal{B}_{n+1}) = \mathbb{P}_{[\omega]_n}^{[\theta^n\omega]_1} f_n \circ T_{[\omega]_{n+1}} + \mathbb{P}_{[\omega]_n}^{[\theta^n\omega]_1} h_n \circ T_{[\omega]_{n+1}} - h_{n+1} \circ T_{[\omega]_{n+1}} = 0.$$

So $(U_n)_{n\in\mathbb{N}_0}$ is a sequence of square integrable reverse martingale differences. Let

$$\sigma_n^2 := \sum_{k=0}^{n-1} \mathbb{E}_{\mu_\omega} U_k^2 = \mathbb{E}_{\mu_\omega}\bigg(\sum_{k=0}^{n-1} U_k\bigg)^2.$$

We check the conditions of Theorem 8.2. Note that $\mathbb{E}$ in the rest of the proof stands for $\mathbb{E}_{\mu_\omega}$.

First we show $\sigma_n^2 \to \infty$ and $\sup_n \mathbb{E}U_n^2 < \infty$. It follows from

$$|\sigma_n - s_n| = \bigg|\mathbb{E}^{1/2}\bigg(\sum_{k=0}^{n-1} U_k\bigg)^2 - \mathbb{E}^{1/2}\bigg(\sum_{k=0}^{n-1} f_k \circ T_{[\omega]_k}\bigg)^2\bigg|$$
$$\leq \mathbb{E}^{1/2}\bigg(\sum_{k=0}^{n-1} U_k - \sum_{k=0}^{n-1} f_k \circ T_0^k\bigg)^2 = \mathbb{E}^{1/2}(h_n \circ T_{[\omega]_n})^2$$
$$\ll \|f\|$$

that $|\sigma_n - s_n|$ is uniformly bounded. So $s_n^2 \to \infty$ implies that $\sigma_n^2 \to \infty$. Since $\|U_n\|_\infty$ is uniformly bounded, $\sup_n \mathbb{E}U_n^2 < \infty$.

Next we show that

$$\sum_{k=0}^{n-1}(\mathbb{E}(U_k^2|\mathcal{B}_{k+1}) - \mathbb{E}(U_k^2)) = o(\sigma_n^2) \quad \mu_\omega\text{-almost surely.}$$

Let $u_n = f_n + h_n - h_{n+1} \circ T_{[\theta^n\omega]_1}$ and let $\tilde{u}_n = u_n^2 - \mathbb{E}U_n^2$. Then $\|\tilde{u}_n\|_\infty \ll \|f\|^2$. Moreover, the Hölder coefficient of $\tilde{u}_n$ is also uniformly bounded because, denoting $[\theta^{n-1}\omega]_1 = i \in \mathcal{W}$,

$$D_\alpha(h_n \circ T_i) = \sup_{x\neq y\in X} \frac{|h_n \circ T_i(x) - h_n \circ T_i(y)|}{d(x, y)^\alpha}$$
$$\leq D_\alpha(h_n) \cdot \sup_{0<d(x,y)<a} \bigg(\frac{d(T_i(x), T_i(y))}{d(x, y)}\bigg)^\alpha + 2a^{-\alpha}\|h_n\|_\infty,$$

https://doi.org/10.1017/etds.2022.60 Published online by Cambridge University Press

which is uniformly bounded by assumption. Let

$$F_n = \sigma_n^{-2} \sum_{k=0}^{n-1} \mathbb{E}(U_k^2 | \mathcal{B}_{k+1}),$$

then

$$\sum_{k=0}^{n-1} (\mathbb{E}(U_k^2 | \mathcal{B}_{k+1}) - \mathbb{E}(U_k^2)) = \sum_{k=0}^{n-1} \mathbb{P}_{[\omega]_k}^{[\theta^k \omega]_1} \tilde{u}_k \circ T_{[\omega]_{k+1}} = \sigma_n^2 (F_n - 1).$$

Applying Proposition 6.1, we have

$$\begin{aligned}
&\mathbb{E}\bigg(\sum_{k=0}^{n-1} \mathbb{P}_{[\omega]_k}^{[\theta^k \omega]_1} \tilde{u}_k \circ T_{[\omega]_{k+1}} \bigg)^2 \\
&\quad \ll \sum_{0 \leq k \leq l \leq n-1} \mathbb{E}(\mathbb{P}_{[\omega]_k}^{[\theta^k \omega]_1} \tilde{u}_k \circ T_{[\omega]_{k+1}} \cdot \mathbb{P}_{[\omega]_l}^{[\theta^l \omega]_1} \tilde{u}_l \circ T_{[\omega]_{l+1}}) \\
&\quad = \sum_{0 \leq k \leq l \leq n-1} \int \mathbb{P}_{[\omega]_k}^{[\theta^k \omega]_{l-k+1}} \tilde{u}_k \cdot \mathbb{P}_{[\omega]_l}^{[\theta^l \omega]_1} \tilde{u}_l \; d\mu_{[\omega]_{l+1}, \theta^{l+1}\omega} \\
&\quad \ll \sum_{l-k+1 \geq k_0} s^{l-k+1} \overline{D} \tilde{u}_k \cdot \mathbb{E} U_l^2 + \sum_{l-k+1 < k_0} \|\tilde{u}_k\|_\infty \cdot \mathbb{E} U_l^2 \\
&\quad \ll k_0 \cdot \sum_{l=0}^{k_0-2} \mathbb{E} U_l^2 + (s^{k_0} + k_0) \cdot \sum_{l=k_0-1}^{n-1} \mathbb{E} U_l^2,
\end{aligned}$$

where in the last inequality, we have used that $\|\tilde{u}_k\|$ is uniformly bounded. Therefore,

$$\mathbb{E}(F_n - 1)^2 = \sigma_n^{-4} \mathbb{E}\bigg(\sum_{k=0}^{n-1} \mathbb{P}_{[\omega]_k}^{[\theta^k \omega]_1} \tilde{u}_k \circ T_{[\omega]_{k+1}} \bigg)^2 \ll \sigma_n^{-4} \sum_{l=0}^{n-1} \mathbb{E} U_l^2 = \sigma_n^{-2}.$$

As $\sigma_n \to \infty$, $\mathbb{E}(F_n - 1)^2 \to 0$. We need to show that it is almost sure convergence. Let $C = \sup_n \mathbb{E} U_n^2$ and let $k_n = \inf\{k : \sigma_k^2 \geq n^2 C\}$. Then $k_n < \infty$, $k_n \to \infty$ and

$$n^2 C \leq \sigma_{k_n}^2 \leq (n^2 + 1) C.$$

Since

$$\sum_n \mathbb{E}(F_{k_n} - 1)^2 \ll \sum_n \sigma_{k_n}^{-2} < \infty,$$

$F_{k_n} \to 1$ almost surely by the Borel–Cantelli lemma. Let $m = m(n) \to \infty$ be such that $k_m \leq n \leq k_{m+1}$, then

$$F_{k_m} \frac{m^2}{(m+1)^2 + 1} \leq F_{k_m} \frac{\sigma_{k_m}^2}{\sigma_{k_{m+1}}^2} \leq F_n \leq F_{k_{m+1}} \frac{\sigma_{k_{m+1}}^2}{\sigma_{k_m}^2} \leq F_{k_{m+1}} \frac{(m+1)^2 + 1}{m^2}.$$

Hence, $F_n \to 1$ almost surely. Lastly, $\sum_n \sigma_n^{-2} \mathbb{E} U_n^2 < \infty$ because $\|U_n\|_\infty$ is uniformly bounded, $|\sigma_n - s_n| \ll \|f\|$ and $\sum_n s_n^{-4} < \infty$ by assumption.

https://doi.org/10.1017/etds.2022.60 Published online by Cambridge University Press

Now we can use Theorem 8.2 to find a sequence of independent centred Gaussian variables $\{Z_k\}$ with $\mathbb{E}Z_k^2 = \mathbb{E}U_k^2$ such that

$$\sup_{0\le k\le n-1}\left|\sum_{i=0}^{k} U_i - \sum_{i=0}^{k} Z_i\right| = o\left(\sqrt{\sigma_n^2 \log\log \sigma_n^2}\right) \quad \text{almost surely.}$$

Since $|\sum_{i=0}^{k} f_i \circ T_{[\omega]_i} - \sum_{i=0}^{k} U_i|$ and $|\sigma_n - s_n|$ are both uniformly bounded, the statement of the theorem follows. □

Remark 8.4. One can verify condition (8.1) on total variance s_n by verifying the inequality

$$\liminf_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1}\mathbb{E}_{\mu_\omega}(f_k^2\circ T_{[\omega]_k}) > 2\sup_{k,m\in\mathbb{N}_0}\left|\sum_{l=k+1}^{k+m}\mathbb{E}_{\mu_\omega}(f_k\circ T_{[\omega]_k}\cdot f_l\circ T_{[\omega]_l})\right|.$$

Assuming that the Ruelle expanding semigroup $\mathcal{S}$ and the potentials φ_i satisfy the conditions of Theorem 5.1, *a priori* the left-hand side of this inequality is positive and the right-hand side is finite for every $f \in \mathcal{H}_\alpha$. A more explicit sufficient condition for f under which this inequality (and equation (8.1)) holds is yet unknown to us.

In that regard, it is also worth noting that the applications of Theorem 2.3 in [**10**] (cf. Theorem 8.2) by Cuny and Merlevède to the iteration of a single, weakly expanding map give rise to explicit function spaces and stronger rates of approximation. However, their results rely on a moderate deviation result for stationary Markov chains by Wu and Zhao in [**36**], which seems not to be available for inhomogeneous Markov chains. Moreover, Dragičević and Hafouta [**14**] and Hafouta [**16**] obtained a vector valued almost sure invariance principle for the sequential iteration of non-uniformly expanding maps. There, the authors obtain a better rate of approximation by assuming an abstract condition on the characteristic functions of the associated process. Finally, we also would like to mention the almost sure invariance principle in [**32**]. There, it was possible to determine an explicit class of functions and sometimes their asymptotic variance such that the almost sure invariance principle holds with respect to sequential systems associated with the continued fraction expansion.

9. *Applications*

In this section, we illustrate some possible applications of our main results, both for conformal iterated function systems and the thermodynamic formalism of free semigroup actions by expanding maps.

9.1. *Non-autonomous conformal iterated function systems.*

The class of non-autonomous conformal iterated function system was introduced and studied in [**27**], and is defined as follows.

Definition 9.1. We refer to $\{X, (\Phi_i : 1 \le i \le k)\}$ as a *non-autonomous conformal iterated function system* if X is a convex, compact subset of $\mathbb{R}^d$ for some $d \in \mathbb{N}$ with $\overline{\mathrm{int}(X)} = X$, and (Φ_i) is a collection $\{\varphi_{i,1}, \ldots, \varphi_{i,k(i)}\}$ of maps from X to X such that:

(1) the following *conformality condition* holds—there exists an open connected set $V \supset X$ such that each $\varphi_{i,j}$ extends to a continuously differentiable conformal diffeomorphism from V into V;
(2) the *open set condition* holds—$\varphi_{i,j}(\text{int}(X)) \cap \varphi_{i,\tilde{j}}(\text{int}(X)) = \emptyset$, for all $1 \leq j < \tilde{j} \leq k(i)$ and $i = 1, \ldots k$;
(3) the following conditions on *bounded distortion and uniform contraction* hold—there exist constants $K \geq 1$ and $\eta \in (0, 1)$ such that for any $n \in \mathbb{N}$ and any choice $(i_1, j_1), \ldots, (i_n, j_n)$, with $i_l \in \{1, \ldots, k\}$ and $1 \leq j_l \leq k(l)$ and all $x, y \in X$, for $\varphi := \varphi_{i_n,j_n} \circ \cdots \circ \varphi_{i_1,j_1}$, we have that

$$\|D\varphi(x)\| \leq K\|D\varphi(y)\|, \quad \|D\varphi(x)\| \leq K\eta^n.$$

As X is assumed to be compact and $k(i) < \infty$ for all $i = 1, \ldots k$, it follows for any compact set $A \subset K$ that $\Phi_i(A) := \bigcup_{j=1}^{k(i)} \varphi_{i,j}(A)$ is compact. Hence, for a given $\omega \in \Sigma$, where $\Sigma = \{(\omega_1\omega_2 \ldots) : 1 \leq \omega_i \leq k\}$, $(\Phi_{\omega_1} \circ \cdots \circ \Phi_{\omega_n}(X))_n$ is a decreasing sequence of compact sets which then implies that the *limit set* J_ω, defined by

$$J_\omega := \lim_{n\to\infty} \Phi_{\omega_1} \circ \Phi_{\omega_2} \circ \cdots \circ \Phi_{\omega_n}(X),$$

is non-empty and compact.

We now derive an averaged version of Bowen's formula to have access to the Hausdorff dimension of these limit sets. To do so, we have to adapt the semigroup setting to the intuitionistic fuzzy set (IFS). First observe that equation (1) in Definition 9.1 implies that $\varphi := \varphi_{i_n,j_n} \circ \cdots \circ \varphi_{i_1,j_1}$ is a well-defined conformal diffeomorphism for any $n \in \mathbb{N}$ and $(i_1, j_1), \ldots, (i_n, j_n)$, with $i_l \in \{1, \ldots, k\}$ and $1 \leq j_l \leq k(l)$. Furthermore, by equation (3), φ is a contraction with rate $K\eta^n$ and, by a standard argument, $x \mapsto \log\|D\varphi(x)\|$ is Lipschitz continuous with respect to a uniform constant.

For $\delta \geq 0$, we now consider the operators, for $w = (\omega_1 \ldots \omega_n)$,

$$\begin{aligned} L^\delta_{\omega_i}(f) &:= \sum_{j=1}^{k(\omega_i)} \|D\varphi_{\omega_i,j}(\,\cdot\,)\|^\delta f \circ \varphi_{\omega_i,j}, \\ L^\delta_w(f) &:= \sum_{j_1,\ldots,j_n} \|D(\varphi_{\omega_1,j_1} \cdots \varphi_{\omega_n,j_n})(\,\cdot\,)\|^\delta f \circ \varphi_{\omega_1,j_1} \cdots \varphi_{\omega_n,j_n} \\ &= L^\delta_{\omega_1} \circ L^\delta_{\omega_2} \circ \cdots \circ L^\delta_{\omega_n}(f) \end{aligned}$$

for f in a suitable function space (the last equality follows from conformality). Now assume that ρ is a probability measure on Σ which satisfies the conditions of Theorem 7.3, that is, $\log d\rho/d\rho \circ \sigma$ is Hölder continuous and the support of ρ is a topological mixing SFT, and, for $n \in \mathbb{N}$,

$$\mathcal{A}^\delta_n := \sum_{w\in\{1,\ldots k\}^n} \rho([w])L^\delta_w.$$

Here $[w]$ represents the cylinder set $\{\omega \in \Sigma : [\omega]_n = w\}$. Observe that the arguments in the proofs of Theorems A and C apply straightforwardly in this context through an interpretation of $\varphi_{\omega_1,j_1} \cdots \varphi_{\omega_n,j_n}$ as an inverse branch of an expanding map. Hence, we obtain uniform and exponential convergence of L^δ_w as $|w| \to \infty$ and of $\mathcal{A}^\delta_n$ as $n \to \infty$.

https://doi.org/10.1017/etds.2022.60 Published online by Cambridge University Press

In particular, for each $\delta \geq 0$, there exists λ_δ such that $\mathcal{A}_n^\delta(\mathbf{1}) \asymp \lambda_\delta^n$. Thus, the annealed pressure function $P : [0, \infty) \to \mathbb{R}$ given by

$$P(\delta) := \lim_{n\to\infty} \frac{1}{n} \log \mathcal{A}_n^\delta(\mathbf{1}) = \log \lambda_\delta$$

is well defined.

LEMMA 9.1. *The function P is continuous and strictly decreasing. Furthermore, $\lim_{\delta\to+\infty} P(\delta) = -\infty$ and $P_0 = \log \lambda_0 \geq \log(\min_i k(i))$, where λ_0 is the spectral radius of the operator defined by*

$$\iota(f) = \sum_{i=1}^{k} k(i) \frac{d\rho}{d\rho \circ \sigma}(i\,\cdot\,) f(i\,\cdot\,).$$

Proof. It follows from the definition and the finiteness of the generating IFS that there exist $\eta_+, \eta_- \in (0, 1)$ such that $\eta_-^n \ll \|D(\varphi_{\omega_1,j_1} \cdots \varphi_{\omega_n,j_n})\| \ll \eta_+^n$. Hence, for $\epsilon > 0$, we have that

$$\eta_-^{n\epsilon} \mathcal{A}_n^\delta(\mathbf{1}) \ll \mathcal{A}_n^{\delta+\epsilon}(\mathbf{1}) \ll \eta_+^{n\epsilon} \mathcal{A}_n^\delta(\mathbf{1}),$$

which implies that $\epsilon \log \eta_- \leq P(\delta+\epsilon) - P(\delta) \leq \epsilon \log \eta_+$. Hence, P is continuous and strictly decreasing. To determine $\lim_{\delta\to+\infty} P(\delta) = -\infty$, observe that

$$\begin{aligned}
\mathcal{A}_{m+n}^\delta(\mathbf{1})(x) &\leq \sum_{|v|=m} \sum_{|w|=n} \rho([vw]) L_v^\delta \circ L_w^\delta(\mathbf{1})(x) \\
&\leq \sum_{|v|=m} \rho([v]) L_v^\delta \bigg(\sum_{|w|=n} \frac{\rho([vw])}{\rho([v])\rho([w])} \rho([w]) L_w^\delta(\mathbf{1}) \bigg)(x) \\
&\leq C \mathcal{A}_m^\delta \circ \mathcal{A}_n^\delta(\mathbf{1})(x), \quad \text{for all } m, n \geq 1
\end{aligned}$$

as there is a uniform bound C for $\rho([v])\rho([w])/\rho([vw])$ by bounded distortion of ρ. Hence, for every fixed $n \geq 1$,

$$\lambda_\delta = \lim_l \sqrt[ln]{\mathcal{A}_{ln}^\delta(\mathbf{1})} \leq \sqrt[n]{C \|\mathcal{A}_n^\delta(\mathbf{1})\|_\infty} \xrightarrow{\delta\to+\infty} 0.$$

To determine $P(0)$, we employ Theorem 7.3 as follows. For $\delta = 0$, $L_i(\mathbf{1}) = k(i)\mathbf{1}$. Hence, by the proof of Theorem 7.3, λ_0 is the spectral radius of ι which is bigger than or equal to $\log(\min_i k(i))$. □

As an immediate corollary, it follows that there exists a unique $\delta_0 > 0$ such that $P(\delta_0) = 0$, provided that $P(0) > 0$, e.g. if $\min_i k(i) > 1$.

THEOREM 9.2. *Assume that $P(0) > 0$. Then, for ρ-almost every ω, the Hausdorff dimension $\dim_H(J_\omega)$ of J_ω is equal to the unique root δ_0 of P.*

Proof. Fix $x \in X$. In analogy to the above pressure function, for $\omega = (\omega_i)$, set

$$P_\omega(\delta) := \limsup_{n\to\infty} \frac{1}{n} \log L_{\omega_1\ldots\omega_n}^\delta(\mathbf{1})(x).$$

To prove almost sure convergence, we employ Kingman's subadditive ergodic theorem. To do so, observe that the shift is ρ-ergodic, and that there exists an equivalent invariant probability measure. Set

$$g_n(\omega) := \sup\{\log L^\delta_{\omega_1\ldots\omega_n}(\mathbf{1})(x) : x \in X\}.$$

By construction, $g_{m+n}(\omega) \leq g_m(\omega) + g_n(\sigma^n(\omega))$. As $g_n(\omega) \asymp \log L^\delta_{\omega_1\ldots\omega_n}(\mathbf{1})(x)$, it now follows from Kingman's subadditive ergodic theorem that $P_\omega(\delta)$ exists almost everywhere and in $L^1(\rho)$, that $P_\omega(\delta)$ is almost surely constant and that the lim sup in the definition in fact is a limit. It follows from these observations that $P_\omega(\delta) = P(\delta)$ almost surely, but for δ fixed. However, by the same argument for Lipschitz continuity of P in the proof above, one obtains that the maps P_ω are equi-Lipschitz continuous. Hence, by choosing a countable and dense set $\{\delta_i\}$, one obtains a set of full measure Ω such that $P_\omega(\delta) = P(\delta)$ for all $\omega \in \Omega$ and $\delta \geq 0$.

We now show that $\dim_H(J_\omega) = \delta_0$ for each $\omega = (\omega_i) \in \Omega$. To do so, we first recall some consequences of conformality. As $\varphi := \varphi_{\omega_1,j_1} \cdots \varphi_{\omega_n,j_n}$ is conformal, it follows that the diameter $\operatorname{diam}(\varphi(X))$ satisfies $\operatorname{diam}(\varphi(X)) \asymp \|D\varphi\| \cdot \operatorname{diam}(X)$. Furthermore, covers by sets of type $\varphi(X)$ are optimal in the following sense. By Lemma 2.7 in [24], or from the proof of Theorem 3.2 in [27], there exists $M \in \mathbb{N}$ such that for each ball B of radius $r > 0$, there exist a subset $W(B)$ of $\{((\omega_1, j_1), \cdots (\omega_n, j_n)) : n \in \mathbb{N}, 1 \leq j_i \leq k(i)\}$ of at most M elements such that:

(1) the elements of $\{\varphi_{\omega_1,j_1} \cdots \varphi_{\omega_n,j_n}(\operatorname{int}(X)) : ((\omega_1, j_1), \ldots (\omega_n, j_n)) \in W(B)\}$ are pairwise disjoint;

(2) $\operatorname{diam}(\varphi_{\omega_1,j_1} \cdots \varphi_{\omega_n,j_n}(X)) \asymp \operatorname{diam}(B)$ for $((\omega_1, j_1), \ldots (\omega_n, j_n)) \in W(B)$;

(3) $B \cap J_\omega \subset \bigcup_{((\omega_1,j_1),\ldots(\omega_n,j_n))\in W(B)} \varphi_{\omega_1,j_1} \cdots \varphi_{\omega_n,j_n}(X)$.

The result now provides access to the δ-Hausdorff measure of J_ω as follows. Assume that $\mathcal{U}$ is a finite cover of J_ω by closed balls. By replacing each $B \in \mathcal{U}$ by $\{\varphi_{\omega_1,j_1} \cdots \varphi_{\omega_n,j_n}(X) : ((\omega_1, j_1), \ldots (\omega_n, j_n)) \in W(B)\}$, we obtain a further cover $\mathcal{V}$ which satisfies

$$\sum_{B\in\mathcal{U}} \operatorname{diam}(B)^\delta \asymp \sum_{A\in\mathcal{V}} \operatorname{diam}(A)^\delta.$$

Hence, to estimate the right-hand side, we may assume without loss of generality that for each $B \in \mathcal{U}$, there exist (ω_i, j_i) such that $B = \varphi_{\omega_1,j_1} \cdots \varphi_{\omega_n,j_n}(X)$. However, Proposition 6.1 implies that for an arbitrary $x \in \operatorname{int}(X)$,

$$\begin{aligned}\mu_\omega(B) &= \lim_{l\to\infty} \frac{L^\delta_{\omega_{n+1}\ldots\omega_{n+l}} \circ L^\delta_{\omega_1\ldots\omega_n}(\mathbf{1}_B)(x)}{L^\delta_{\omega_1\ldots\omega_{n+l}}(\mathbf{1})(x)} \\ &\asymp \|D\varphi_{\omega_1,j_1} \cdots \varphi_{\omega_n,j_n}\|^\delta \lim_{l\to\infty} \frac{L^\delta_{\omega_{n+1}\ldots\omega_{n+l}}(\mathbf{1})(x)}{L^\delta_{\omega_1\ldots\omega_{n+l}}(\mathbf{1})(x)} \asymp \operatorname{diam}(B)^\delta \lambda^{-1}_{\omega_1\ldots\omega_n,\sigma^n\omega}.\end{aligned}$$

Setting $|B| = n$, this implies that

$$\sum_{B\in\mathcal{U}} \operatorname{diam}(B)^\delta \asymp \sum_{B\in\mathcal{U}} \lambda_{\omega_1\ldots\omega_{|B|},\sigma^{|B|}\omega} \mu_\omega(B).$$

Now assume that the interiors of the elements of $\mathcal{U}$ are disjoint. Then $\sum \mu_\omega(B) = 1$ and the asymptotics of $\sum \operatorname{diam}(B)^\delta$ as $\max \operatorname{diam}(B) \to 0$ are determined by the asymptotics of $\lambda_{\omega_1 \ldots \omega_n, \sigma^n \omega}$ as $n \to \infty$. Hence, if $\delta > \delta_0$, then the δ-Hausdorff measure of J_ω is 0 and if $\delta < \delta_0$, then the δ-Hausdorff measure of J_ω is ∞. This implies that $\dim_H(J_\omega) = \delta_0$. □

9.2. *Thermodynamic formalism of semigroup actions.* In this subsection, we will provide some applications of our results to the setting of finitely generated free semigroup actions.

Let X be a compact metric space, $\varphi : X \to \mathbb{R}$ be a continuous potential and let $G_1 = \{g_1, g_2, \ldots, g_k\}$ be a finite set of continuous self maps on X, for some $k \geq 2$. The semigroup $\mathcal{S}$ generated by G_1 induces a continuous semigroup action given by

$$\begin{aligned} \mathbb{S} : \mathcal{S} \times X &\to X \\ (g, x) &\mapsto g(x), \end{aligned}$$

meaning that for any $\underline{g},\ \underline{h} \in \mathcal{S}$ and every $x \in X$, we have $\mathbb{S}(\underline{g}\,\underline{h}, x) = \mathbb{S}(\underline{g}, \mathbb{S}(\underline{h}, x))$. The thermodynamic formalism of semigroup actions faces several difficulties. On one hand, while probability measures which are invariant by all generators may fail to exist, in opposition to the case of group actions, there are evidences that the stationary measures seem not sufficient to describe the dynamics. On the other hand, the existence of some distinct concepts of topological pressure for group and semigroup actions makes it necessary to test their effectiveness to describe the dynamics. In the case of free semigroup actions, the coding of the dynamics by the full shift suggests to consider the skew-product

$$\begin{aligned} F : \{1, 2, \ldots, k\}^{\mathbb{N}} \times X &\to \{1, 2, \ldots, k\}^{\mathbb{N}} \times X \\ (\omega, x) &\mapsto (\sigma(\omega), g_{\omega_1}(x)). \end{aligned} \tag{9.1}$$

Moreover, a random walk on the semigroup $\mathcal{S}$ can be modelled by a Bernoulli probability measure $\mathbb{P}$ on $\{1, 2, \ldots, k\}^{\mathbb{N}}$. The pressure $P_{\text{top}}(\mathbb{S}, \phi, \mathbb{P})$ of the semigroup action determined by that random walk coincides with the annealed topological pressure $P^{(a)}_{\text{top}}(F, \tilde{\phi}, \mathbb{P})$ of the random dynamical system determined by F, associated to the potential $\tilde{\phi} : \{1, 2, \ldots, k\}^{\mathbb{N}} \times X \to \mathbb{R}$ given by $\tilde{\phi}(\omega, x) = \phi(x)$ (cf. Proposition 4.1 in [7]). In particular, $P_{\text{top}}(\mathbb{S}, \phi, \mathbb{P})$ coincides with the logarithm of the spectral radius of the averaged transfer operator

$$\mathcal{A}_1(f) = \int L_{g_\omega}(f)\, d\mathbb{P}(\omega).$$

Furthermore, if $P_{\text{top}}(\mathbb{S}, 0, \mathbb{P}) < \infty$, then entropy and invariant measures can be defined through a functional analytic approach, which culminates in the variational principle

$$P_{\text{top}}(\mathbb{S}, \phi, \mathbb{P}) = \sup_{\{\nu \in \mathcal{M}(X)\,:\, \Pi(\nu,\sigma) \neq \emptyset\}} \left\{ h_\nu(\mathbb{S}, \mathbb{P}) + \int \phi\, d\nu \right\} \tag{9.2}$$

(we refer the reader to [7] for the definitions and more details). If all generators are Ruelle expanding maps and ϕ is Hölder continuous, then there exists a unique equilibrium state for the semigroup action $\mathbb{S}$ with respect to ϕ and this can be characterized either as a marginal of the unique equilibrium state for the annealed random dynamics or as the

https://doi.org/10.1017/etds.2022.60 Published online by Cambridge University Press

unique probability on X obtained as the limit of the equidistribution along pre-orbits associated to the semigroup dynamics by

$$e^{-nP_{\mathrm{top}}(\mathbb{S},\phi,\mathbb{P})}\mathcal{A}_1^{*n}\delta_x = e^{-nP_{\mathrm{top}}(\mathbb{S},\phi,\mathbb{P})}\int_{\mathcal{W}_n}\Bigg[\sum_{g_\omega(y)=x}\delta_y\Bigg]\,d\mathbb{P}(\omega)$$

(we refer the reader to [6, §9] and [7, Theorem B] for more details). A more general formulation, considering more general probabilities on semigroup actions rather than random walks, was not available up to now as the thermodynamic formalism of the associated annealed dynamics needed to be described through a sequence of transfer operators instead of a single averaged operator.

Our results allow not only to consider the thermodynamic formalism of semigroup actions with respect to more general probabilities in the base, but also to provide important asymptotic information on the convergence to equilibrium states. Indeed, in general, if one endows the semigroup $\mathcal{S}$ with a probability generated by a Markov measure $\mathbb{P}$ on $\{1, 2, \ldots, k\}^{\mathbb{N}}$, then it is natural to define the topological pressure of the semigroup action $\mathbb{S}$ by

$$P_{\mathrm{top}}(\mathbb{S}, \phi, \mathbb{P}) = \limsup_{n\to\infty}\frac{1}{n}\log\|\mathcal{A}_n(1)\|_\infty \tag{9.3}$$

where, as before, $\mathcal{A}_n(f) = \int_{\omega\in\mathcal{W}_n} L_{g_{\omega_1\omega_2\ldots\omega_n}}(f)\,d\mathbb{P}(\omega)$ (compare to the definition of topological pressure of a semigroup action in [7, §2.6]). Our main results have the following immediate consequences.

COROLLARY 9.3. *Given $x \in X$, the sequence of probability measures on X defined as*

$$\nu_n^x := \frac{\mathcal{A}_n^*(\delta_x)}{\mathcal{A}_n(\mathbf{1})(x)}, \quad n \geq 1$$

is weak convergent to some probability $\nu = h\,d\pi$ on X (independently of x). Moreover, the convergence is exponentially fast with respect to the Wasserstein distance.*

9.3. *A boundary of equilibria.* As in the section before, we now assume that X is compact and that there is only one potential $\varphi : X \to \mathbb{R}$. However, in contrast to the approach via the free semigroup, we are now interested in identifying elements in the semigroup $\mathcal{S}$ which are dynamically close and use this information to define a compactification of the discrete set $\mathcal{S}$. However, as the topology will rely on the associated equilibrium states, we have to extend the semigroup by considering also the potential function. That is, for $\mathbb{G}_1 := \{(g_1, \varphi), (g_2, \varphi), \ldots (g_k, \varphi)\}$, we consider

$$\mathbb{G} := \{(g, \psi) : \text{there exists } n \in \mathbb{N},\ j_1, \ldots, j_n \text{ such that } (g, \psi) = (g_{i_1}, \varphi) * \cdots * (g_{i_n}, \varphi)\},$$

where

$$(g_1, \psi_1) * (g_2, \psi_2) := (g_1 \circ g_2, \psi_2 + \psi_1 \circ g_2)$$

is also the product on $\mathbb{G}$.

As a first step, we begin with the definition of a metric on the countable set $\mathcal{W}^* := \{w : |w| < \infty\}$ of finite words. For finite words $v = (v_1 \ldots v_m)$ and $w = (w_1 \ldots w_n)$ in $\mathcal{W}^*$,

https://doi.org/10.1017/etds.2022.60 Published online by Cambridge University Press

set $d_{\mathcal{W}^*}(v, w) = 0$ for $v = w$ and

$$d_{\mathcal{W}^*}(v, w) := 2^{-\min\{k : v_k \neq w_k \text{ or } k > \min\{m,n\}\}} + 2^{-\min\{k : v_{m+1-k} \neq w_{n+1-k} \text{ or } k > \min\{m,n\}\}},$$

for $v \neq w$. Observe that $d_{\mathcal{W}^*}$ is a metric, that $\mathcal{W}^*$ is discrete with respect to this metric and that two words are close if they have the same beginning and ending. In particular, Cauchy sequences either have to be eventually constant or have to grow from the interior of a word. The reason for this construction is based on the following observation. Let $\underline{w}$ and $\overline{w}$ refer to the periodic extensions of w to the left and the right, respectively, as defined in Remark 6.4. Then, by Proposition 6.3, the map $w \to \mu_{\underline{w},\overline{w}}$ is Hölder continuous with respect to $d_{\mathcal{W}^*}$. In particular, $d_{\mathcal{W}^*}$ can be seen as a metric on the free semigroup which is compatible with the Wasserstein distance of the associated equilibrium states.

Second, we define a metric on $\mathbb{G}$ which does not depend on the choice of $w \in \mathcal{W}^*$ for the representation of $(g, \psi) = (T_w, \varphi_w)$. To do so, define for $g \in \mathcal{S}$,

$$\kappa(g) := \lim_{\epsilon \to 0} \inf \left\{ \frac{d(g(x), g(y))}{d(x, y)} : 0 < d(x, y) < \epsilon \right\},$$

and note that as the semigroup is Ruelle expanding with parameter $\lambda \in (0, 1)$, we have that $\kappa(T_w) \geq \lambda^{-|w|}$. Furthermore, for $(g, \psi) \in \mathbb{G}$, let $\mu_{g,\psi}$ be the unique equilibrium state for the potential ψ and the map g, that is, if $(g, \psi) = (T_w, \varphi_w)$, then $\mu_{g,\psi} = \mu_{\underline{w},\overline{w}}$. Now set

$$d_{\mathbb{G}}((g, \psi_1), (h, \psi_2)) := \begin{cases} \overline{W}(\mu_{g,\psi_1}, \mu_{h,\psi_2}) + \dfrac{1}{\kappa(g)} + \dfrac{1}{\kappa(h)} & , (g, \psi_1) \neq (h, \psi_2), \\ 0 & , (g, \psi_1) = (h, \psi_2). \end{cases}$$

The following proposition summarizes the basic topological facts. The proof is omitted as the assertions almost immediately follow from the definitions and Proposition 6.3.

PROPOSITION 9.4. *Assume that $g_1, \ldots, g_k$ are Ruelle expanding and jointly topological mixing, and that φ is Hölder continuous. Then, for the objects defined above, the following hold.*

(1) *$(\mathcal{W}^*, d_{\mathcal{W}^*})$ and $(\mathbb{G}, d_{\mathbb{G}})$ are discrete, metric spaces.*
(2) *The map $w \mapsto (T_w, \varphi_w)$ is Hölder continuous.*
(3) *A sequence $((g_n, \psi_n))_n$ in $\mathbb{G}$ is a Cauchy sequence if and only if $\kappa(g_n) \to \infty$ and (μ_{g_n,ψ_n}) converges in the weak*-topology. Moreover, two Cauchy sequences have the same limit if and only if their sequences of equilibrium states have the same limit.*
(4) *For the boundary $\partial\mathbb{G}$ of the completion with respect to $d_{\mathbb{G}}$, identified with limits of Cauchy sequences $((g_n, \psi_n))_n$ in $\mathbb{G}$, we have that the map*

$$\partial\mathbb{G} \to \{\mu_{\sigma,\omega} : \sigma \in \Sigma^-, \omega \in \Sigma\}, \ ((g_n, \psi_n))_n \mapsto \lim_{n\to\infty} \mu_{g_n,\psi_n}$$

is Lipschitz continuous and onto.

Observe that the result provides a description of $\partial\mathbb{G}$ as a set of equivalence classes of Cauchy sequences, that is, two sequences are considered to be equivalent if they have the same limit. However, it seems to be impossible to obtain an explicit description of $\partial\mathbb{G}$ in general. We close with two examples where this is possible. In the first example, $\partial\mathbb{G}$ is trivial whereas in the second example, $\partial\mathbb{G}$ is equal to Σ^-.

PROPOSITION 9.5. *If $\mathbb{G}$ is Abelian, then $\partial\mathbb{G}$ is a point.*

Proof. Assume that $(g_1, \psi_1), (g_2, \psi_2) \in \mathbb{G}$, and denote by L_i the corresponding Ruelle operators. As $\mathbb{G}$ is Abelian, it immediately follows that $L_1 L_2 = L_2 L_1$. Now assume that the h_i are the unique positive Hölder functions (up to colinearity) and $\lambda_i > 0$ such that $L_i(h_i) = \lambda_i h_i$, given by Ruelle's theorem. Hence, $L_2(L_1(h_2)) = L_1(L_2(h_2)) = \lambda_2 L_1(h_2)$. As $L_1(h_2)$ is positive, it follows that $L_1(h_2)$ and h_1 are colinear, that is, $L_1(h_2)$ is a multiple of h_1 and $\lambda_1 = \lambda_2$. The same argument then shows that the L_i^*-eigenmeasures coincide. Hence, after normalizing, we obtain that $\mu_{g_1,\psi_1} = \mu_{g_2,\psi_2}$. In particular, $\{\mu_{\sigma,\omega} : \sigma \in \Sigma^-, \omega \in \Sigma\}$ is a singleton. □

Example 9.6. Let $T : [0, 1] \to [0, 1]$, $x \mapsto 4x(\mathrm{mod} 1)$ and $S = U^{-1}TU$, where

$$U : [0, 1] \to [0, 1], \quad x \mapsto \begin{cases} 3x/2, & 0 \le x < 1/8, \\ x + 1/16, & 1/8 \le x < 3/8, \\ x/2 + 1/4, & 3/8 \le x < 1/2, \\ x, & 1/2 < x \le 1. \end{cases}$$

PROPOSITION 9.7. *The semigroup $\mathcal{S}$ generated by $\{S, T\}$ is a free semigroup, that is, two elements in $\mathcal{S}$ coincide if and only if they have the same representation as a product of the generators. Moreover, $\partial\mathbb{G} \cong \Sigma^-$, where $\mathbb{G}$ is the semigroup generated by $(T, 0)$ and $(S, 0)$.*

Proof. The proof relies on the construction of a family of renormalization operators acting on the set of orientation-preserving homeomorphisms f in such a way that

$$T^n \circ \Xi_n(f) = f \circ T^n,$$

as this allows to associate to each element $g = S^{m_k}T^{n_k} \cdots S^{m_1}T^{n_1}$ in $\mathcal{S}$ a uniquely determined normal form $T^{m_1+n_1+\cdots m_k+n_k} \circ f_g$, where f_g is an orientation-preserving homeomorphism. The uniqueness of the normal form is a consequence of the choice of U as the compositions with U and U^{-1} act as markers in the following way. For an orientation-preserving homeomorphism f, it is shown below that $\|\Xi^n(f) - \mathrm{id}\|_\infty = 4^{-n}\|f - \mathrm{id}\|_\infty$, and that the composition $\Xi_n(f) \circ U^{\pm 1}$ leaves invariant the right half of $\Xi_n(f)$, whereas the left half is marked by a positive or negative bump of size bigger than $\|\Xi^n(f) - \mathrm{id}\|_\infty$.

Construction and properties of Ξ_n. Let $f : [0, 1] \to [0, 1]$ be a homeomorphism which fixes 0 and 1 and define for $x \in [k/4^n, (k + 1)/4^n]$,

$$\Xi_n(f)(x) := (T^n|_{[k/4^n,(k+1)/4^n]})^{-1} \circ f \circ T^n(x) = 4^{-n}(f(4^n x - k) + k).$$

Then, as it can be easily seen, $T^n \circ \Xi_n(f) = f \circ T^n$ and $\Xi_n(f)(k/4^n) = k/4^n$ for all $k = 0, \ldots, 4^n$. In particular, as $\Xi_n(f)|_{[k/4^n,(k+1)/4^n]}$ is a homeomorphism, $\Xi_n(f)$ is a

https://doi.org/10.1017/etds.2022.60 Published online by Cambridge University Press

homeomorphism. Moreover, for $x \in [k/4^n, (k+1)/4^n]$, we have

$$\begin{aligned}\Xi_n(f)(x) - x &= 4^{-n}(f(4^n x - k) + k) - x \\ &= 4^{-n}(f(4^n x - k) - (4^n x - k)) = 4^{-n}(f \circ T^n(x) - T^n(x)).\end{aligned}$$

That is, Ξ_n contracts the distance to the identity by the factor 4^{-n}. We now proceed with an analysis of the concatenations $\Xi_n(f) \circ U$ and $\Xi_n(f) \circ U^{-1}$, where f is a homeomorphism with $\|f - \mathrm{id}\|_\infty \leq 1/12$. First note that

$$U(x) - x = \begin{cases} x/2, & x \in [0, \frac{1}{8}), \\ 1/16, & x \in [\frac{1}{8}, \frac{3}{8}), \\ -x/2 + 1/4, & x \in [\frac{3}{8}, \frac{1}{2}), \\ 0, & x \in [\frac{1}{2}, 1], \end{cases} \quad U^{-1}(x) - x = \begin{cases} -x/3, & x \in [0, \frac{3}{16}), \\ -1/16, & x \in [\frac{3}{16}, \frac{7}{16}), \\ x - 1/2, & x \in [\frac{7}{16}, \frac{1}{2}), \\ 0, & x \in [\frac{1}{2}, 1], \end{cases}$$

and observe that, by construction, $\Xi_n(f) - \mathrm{id}$ is periodic with period 4^{-n}. However, as $[\frac{1}{8}, \frac{3}{8})$, $[3/16, 7/16)$ and $[\frac{1}{2}, 1]$ are all of length bigger than or equal to $1/4$, we obtain that

$$\begin{aligned}\max_{x\in[0,1]} (\Xi_n(f)(U(x)) - x) &= \max_{x\in[1/8,3/8)} (\Xi_n(f)(U(x)) - U(x) + U(x) - x) \\ &= 4^{-n} \max_{x\in[0,1]} (f(x) - x) + \frac{1}{16} = \frac{1}{4^n \cdot 12} + \frac{1}{16} \leq \frac{1}{12},\end{aligned}$$

and, repeating the argument, $\|\Xi_n(f) \circ U^j - \mathrm{id}\|_\infty \leq 1/12$, for $j = \pm 1$.

In other words, the space $\mathfrak{H}$ of orientation-preserving homeomorphisms with $\|f - \mathrm{id}\|_\infty \leq 1/12$ is invariant under the operation $f \mapsto \Xi_n(f) \circ U^j$. Moreover, we have that

$$\|\Xi_n(f) \circ U^j - U^j\|_\infty = 4^{-n}\|\Xi_n(f) - \mathrm{id}\|_\infty = 4^{-n}\|f - \mathrm{id}\|_\infty \leq \tfrac{1}{48}. \tag{9.4}$$

Coding of $\mathbb{G}$. Assume that $g = S^{m_k}T^{n_k} \cdots S^{m_1}T^{n_1}$ for some $k \in \mathbb{N}$ and $m_i, n_i \in \mathbb{N} \cup \{0\}$. As $U, U^{-1} \in \mathfrak{H}$, it follows from an iterated application of $\Xi_n(\cdot) \circ U^j$ that there exists a homeomorphism $f_g \in \mathfrak{H}$ such that $g = T^n \circ f_g$, where $n = \sum_{i=1}^k m_i + n_i$. Moreover, as T^n is a local homeomorphism, $f = f_g$ is uniquely determined.

Now assume that $g = S^{m_k}T^{n_k} \cdots S^{m_1}T^{n_1} \in \mathcal{S}$ where, without loss of generality, $m_1, \ldots, m_{k-1} \neq 0$ and $n_2, \ldots, n_k \neq 0$. We now show how to determine m_1 and n_1 from f in a unique way.

Case 1. If $m_1 = 0$, then $k = 1$, $g = T^{n_1}$ and $f = \mathrm{id}$.

Case 2. If $m_1 \neq 0$ and $n_1 \neq 0$, then $k > 1$ and for $\bar{f} := f_{S^{m_k}T^{n_k} \ldots S^{m_1}}$, we have that $f = \Xi_{n_1}(\bar{f})$. It now follows from equation (9.4) that $\bar{f} - \mathrm{id}$ is strictly positive on $[1/8, 3/8]$ and has zeros in $[1/2, 1]$. Therefore, n_1 is determined by the periodicity of $f - \mathrm{id}$, and $\bar{f}(x) = f(2^{n_1})(x)$. The value of m_1 is then determined by applying Case 3 to $S^{m_k}T^{n_k} \cdots S^{m_1}$ and $\bar{f}$.

Case 3. If $m_1 \neq 0$ and $n_1 = 0$, then $k \geq 1$ and for $\bar{f} := f_{S^{m_k}T^{n_k} \ldots T^{m_2}}$, we have that $f = \Xi_{m_1}(\bar{f} \circ U^{-1}) \circ U$ or, equivalently, $f \circ U^{-1} = \Xi_{m_1}(\bar{f})$. Hence, to repeat the above argument based on periodicity, we have to show that the left half of $\bar{f} - \mathrm{id}$ is somehow marked. If $k = 1$, then $\bar{f} = U^{-1}$ and, in particular, $\bar{f}$ is strictly negative on $[3/16, 7/16]$

and has zeros in $[1/2, 1]$. Hence, m_1 can be determined through the period of $f \circ U^{-1}$. However, if $k > 1$, then $n_2 > 0$ and the same argument is applicable as equation (9.4) implies that $\bar{f}$ is strictly negative on $[3/16, 7/16]$ and has zeros in $[1/2, 1]$.

By iterating this procedure, one then recovers $m_2, \ldots, m_k$ and $n_2, \ldots, n_k$ from f. Furthermore, as the m_i and n_i only depend on the period, it follows that the relation between f and these values is one-to-one. This then implies that the map

$$\mathcal{S} \to \{f_g : g \in \mathcal{S}\}, \quad (w_1 \ldots w_n) \mapsto f_{w_n \circ \cdots \circ w_1}$$

is a bijection, and, as an immediate corollary, $\mathcal{S}$ is a free semigroup.

The associated measures of maximal entropy. Now fix a Hölder function h, an element $g \in \mathcal{S}$ and let $n \in \mathbb{N}$ be given by $g = T^n \circ f_g$. Then the Ruelle operators L_g and L_T associated to g and T, respectively, satisfy

$$L_g(h)(x) = \sum_{g(y)=x} h(y) = \sum_{T^n z = x} h(f_g^{-1}(z)) = L_T^n(h \circ f_g^{-1})(x),$$

$$\frac{L_g(hL_g(\mathbf{1}))}{L_{g^2}(\mathbf{1})} = \frac{L_g(4^n h)}{4^{2n}} = \frac{1}{4^n} L_T^n(h \circ f^{-1}).$$

By Proposition 6.3, the measures of maximal entropy μ_g and μ_T of g and T, respectively, satisfy $\overline{W}(\mu_g, \mu_T \circ f_g) \ll s^n$. Hence, $\mu_g = \lim_{l \to \infty} \mu_T \circ f_{g^l}$. However, this result also implies that for an infinite word $(v_i) \in \{S, T\}^{\mathbb{N}}$, the sequence $\mu_{g_{v_l \cdots v_1}}$ is a Cauchy sequence and therefore convergent. It remains to show that the mapping from (v_i) to this limit is injective. To do so, let $(v_i) \neq (w_i)$ be different elements in $\{S, T\}^{\mathbb{N}}$. Then, by applying the construction of the n_i and m_i above to infinite words, it follows that $\mu_{g_{v_l \cdots v_1}} \neq \mu_{g_{w_l \cdots w_1}}$ for all l sufficiently large. Furthermore, it can be deduced from the recursive construction of f_g that there exists an open set A and $\epsilon > 0$ such that $f_{v_l \cdots v_1}(x) - f_{w_l \cdots w_1}(x) > \epsilon$ for all $x \in A$ and all l sufficiently large. Hence, $\lim_l \mu_{g_{v_l \cdots v_1}} \neq \lim_l \mu_{g_{w_l \cdots w_1}}$. □

Acknowledgements. First of all, the authors would like to thank the anonymous referee whose comments helped to improve the exposition of the paper. Furthermore, M.S. was supported by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (PROEX) and Conselho Nacional de Desenvolvimento Científico e Tecnológico (PQ 312632/2018-5, Universal 426814/2016-9). P.V. was partially supported by Centro de Matemática de Universidade do Porto (UID/MAT/00144/2013), funded by Fundação para a Ciência e Tecnologia - Portugal with national (MEC) and European structural funds through the programs FEDER, under the partnership agreement PT2020, and by Fundação para a Ciência e Tecnologia - Portugal, through the grant CEECIND/03721/2017 of the Stimulus of Scientific Employment, Individual Support 2017 Call. X.Z. was supported by Fundação de Amparo à Pesquisa do Estado de S. Paulo grant no. 2018/15088-4.

References

[1] J. Atnip, G. Froyland, C. González-Tokman and S. Vaienti. Thermodynamic formalism for random weighted covering systems. *Comm. Math. Phys.* **386** (2021), 819–902.

[2] V. Baladi. Correlation spectrum of quenched and annealed equilibrium states for random expanding maps. *Comm. Math. Phys.* **186** (1997), 671–700.

[3] M. Bessa and M. Stadlbauer. On the Lyapunov spectrum of relative transfer operators. *Stoch. Dyn.* **16**(6) (2016), 1650024.
[4] T. Bogenschütz and V. M. Gundlach. Ruelle's transfer operator for random subshifts of finite type. *Ergod. Th. & Dynam. Sys.* **15** (1995), 413–447.
[5] X. Bressaud, R. Fernández and A. Galves. Decay of correlations for non-Hölderian dynamics. A coupling approach. *Electron. J. Probab.* **4**(3) (1999), 19 pp (electronic).
[6] M. Carvalho, F. Rodrigues and P. Varandas. Semigroup actions of expanding maps. *J. Stat. Phys.* **116**(1) (2017), 114–136.
[7] M. Carvalho, F. Rodrigues and P. Varandas. A variational principle for free semigroup actions. *Adv. Math.* **334** (2018), 450–487.
[8] A. Castro, F. Rodrigues and P. Varandas. Stability and limit theorems for sequences of uniformly hyperbolic dynamics. *J. Math. Anal. Appl.* **480** (2019), 123426.
[9] J. P. Conze and A. Raugi. Limit theorems for sequential expanding dynamical systems on [0, 1]. *Ergodic Theory and Related Fields (Contemporary Mathematics, 430)*. Ed. I. Assani. American Mathematical Society, Providence, RI, 2007, pp. 89–121.
[10] C. Cuny and F. Merlevède. Strong invariance principles with rate for 'reverse' martingale differences and applications. *J. Theoret. Probab.* **28**(1) (2015), 137–183.
[11] M. Denker and M. Gordin. Gibbs measures for fibred systems. *Adv. Math.* **148**(2) (1999), 161–192.
[12] M. Denker, M. Gordin and S.-M. Heinemann. On the relative variational principle for fibre expanding maps. *Ergod. Th. & Dynam. Sys.* **22**(3) (2002), 757–782.
[13] D. Dragičević, G. Froyland, C. González-Tokman and S Vaienti. Almost sure invariance principle for random piecewise expanding maps. *Nonlinearity* **31**(5) (2018), 2252–2280.
[14] D. Dragičević and Y. Hafouta. Almost sure invariance principle for random dynamical systems via Gouëzel's approach. *Nonlinearity* **34**(10) (2021), 6773–6798.
[15] A. M. Fisher. Small-scale structure via flows. *Fractal Geometry and Stochastics III*. Eds. C. Bandt, U. Mosco and M. Zähle. Birkhäuser Basel, Basel, 2004, pp. 59–78.
[16] Y. Hafouta. A vector valued almost sure invariance principle for time dependent non-uniformly expanding dynamical systems. *Preprint*, 2020, arXiv:1910.12792.
[17] M. Hairer and J. C. Mattingly. Spectral gaps in Wasserstein distances and the 2D stochastic Navier–Stokes equations. *Ann. Probab.* **36**(6) (2008), 2050–2091.
[18] N. Haydn, M. Nicol, A. Török and S. Vaienti. Almost sure invariance principle for sequential and non-stationary dynamical systems. *Trans. Amer. Math. Soc.* **369** (2017), 5293–5316.
[19] L. Heinrich. Mixing properties and central limit theorem for a class of non-identical piecewise monotonic C^2—Transformations. *Math. Nachr.* **181** (1996), 185–214.
[20] J. Jaerisch and H. Sumi. Dynamics of infinitely generated nicely expanding rational semigroups and the inducing method. *Trans. Amer. Math. Soc.* **369**(9) (2017), 6147–6187.
[21] Y. Kifer. Perron–Frobenius theorem, large deviations, and random perturbations in random environments. *Math. Z.* **222**(4) (1996), 677–698.
[22] Y. Kifer. Limit theorems for random transformations and processes in random environments. *Trans. Amer. Math. Soc.* **350**(4) (1998), 1481–1518.
[23] B. R. Kloeckner, A. O. Lopes and M. Stadlbauer. Contraction in the Wasserstein metric for some Markov chains, and applications to the dynamics of expanding maps. *Nonlinearity* **28**(11) (2015), 4117–4137.
[24] R. D. Mauldin and M. Urbański. Dimensions and measures in infinite iterated function systems. *Proc. Lond. Math. Soc. (3)* **73**(1) (1996), 105–154.
[25] R. D. Mauldin and M. Urbański. *Graph Directed Markov Systems: Geometry and Dynamics of Limit Sets (Cambridge Tracts in Mathematics, 148)*. Cambridge University Press, Cambridge, 2003.
[26] V. Mayer, B. Skorulski and M. Urbański. *Distance Expanding Random Mappings, Thermodynamical Formalism, Gibbs Measures and fractal Geometry (Lecture Notes in Mathematics, 2036)*. Springer, Heidelberg, 2011.
[27] L. Rempe-Gillen and M. Urbański. Non-autonomous conformal iterated function systems and Moran-set constructions. *Trans. Amer. Math. Soc.* **368**(3) (2016), 1979–2017.
[28] D. Ruelle. The thermodynamic formalism for expanding maps. *Comm. Math. Phys.* **125**(2) (1989), 239–262.
[29] D. Ruelle. *Thermodynamic Formalism (Cambridge Mathematical Library)*, 2nd edn. Cambridge University Press, Cambridge, 2004.
[30] O. M. Sarig. Existence of Gibbs measures for countable Markov shifts. *Proc. Amer. Math. Soc.* **131**(6) (2003), 1751–1758.
[31] M. Stadlbauer. Coupling methods for random topological Markov chains. *Ergod. Th. & Dynam. Sys.* **37**(3) (2017), 971–994.
[32] M. Stadlbauer and X. Zhang. On the law of the iterated logarithm for continued fractions with sequentially restricted partial quotients. *Nonlinearity* **34** (2021), 1389–1407.

[33] H. Sumi and M. Urbanski. The equilibrium states for semigroups of rational maps. *Monatsh. Math.* **156**(4) (2009), 371–390.

[34] H. Sumi and M. Urbanski. Transversality family of expanding rational semigroups. *Adv. Math.* **234** (2013), 697–734.

[35] M. Viana and K. Oliveira. *Foundations of Ergodic Theory*. Cambridge University Press, Cambridge, 2016.

[36] W. B. Wu and Z. Zhao. Moderate deviations for stationary processes. *Statist. Sinica* **18**(2) (2008), 769–782.