Ergod. Th. & Dynam. Sys. (2017), **37**, 971–994 doi:10.1017/etds.2015.61
© *Cambridge University Press, 2015*

Coupling methods for random topological Markov chains

MANUEL STADLBAUER

Instituto de Matemática, Universidade Federal do Rio de Janeiro, CP 68530, CEP 21945-970, Rio de Janeiro (RJ), Brazil (e-mail: manuel.stadlbauer@gmail.com)

(*Received* 21 *December* 2013 *and accepted in revised form* 5 *June* 2015)

Abstract. We apply coupling techniques in order to prove that the transfer operators associated with random topological Markov chains and non-stationary shift spaces with the big images and preimages property have a spectral gap.

1. *Introduction*

In this note we give a further contribution to thermodynamical formalism for random countable Markov chains by applying coupling techniques and methods from optimal transport in order to circumvent problems arising from the randomized setting. In case of a random topological Markov chain with finitely many states, it was shown by Bogenschütz and Gundlach [**2**] and Kifer [**13**] that the method of convex cones can be adapted to the random setting in order to obtain exponential convergence to the eigenfunction of the Ruelle operator. However, even in case of deterministic topological Markov chains with countably many states, these cone techniques are not applicable and have to be replaced by proving quasi-compactness of Ruelle's operator using the method developed by Doeblin and Fortet [**7**], Tulcea and Marinescu [**20**] and Hennion [**12**], which is, at least so far, unavailable in the setting of random dynamical systems. In fact, basic examples without measurable families of invariant functions as in [**8**, Proposition 1] suggest that there might be no randomized version of the Doeblin–Fortet method. Therefore, we employ coupling techniques from the theory of Markov operators and optimal transport. It is worth noting that this approach does not require a Lasota–Yorke condition (as, e.g., by Buzzi in [**4**]).

In here, the setting with respect to random topological Markov chains follows closely the one in [**18**], which we sketch now, postponing the details to §3. A random topological

https://doi.org/10.1017/etds.2015.61 Published online by Cambridge University Press

Markov chain is a random bundle transformation, that is, a commuting diagram (or fibred system)

$$\begin{array}{ccc} X & \xrightarrow{T} & X \\ {\scriptstyle\pi}\downarrow & & \downarrow{\scriptstyle\pi} \\ \Omega & \xrightarrow{\theta} & \Omega \end{array}$$

where θ is an ergodic automorphism of the abstract probability space (Ω, P) and π is onto and measurable. In case of a random topological Markov chain, X is a subset of $\mathbb{N}^{\mathbb{N}} \times \Omega$ and each fibre $X_\omega := \pi^{-1}(\{\omega\})$ has the following topological Markov structure: for almost every (a.e.) ω, there is an associated alphabet $\mathscr{W}_\omega^1$ and a matrix $A_\omega = (\alpha_{ij}^\omega : i \in \mathscr{W}_\omega^1,\ j \in \mathscr{W}_{\theta\omega}^1)$, called the (random) transition matrix. Then a pair $uv \in \mathscr{W}_\omega^1 \times \mathscr{W}_{\theta\omega}^1$ is called *ω-admissible* if $\alpha_{uv}^\omega = 1$. Moreover, $(x_0, x_1, x_2, \ldots) \in X_\omega$ if and only if $x_i \in \mathscr{W}_{\theta^i\omega}^1$ and $x_i x_{i+1}$ is $\theta^i\omega$-admissible for all $i = 0, 1, \ldots$. Note that X_ω comes with a natural, non-random topology defined through the shift metric on sequence spaces as defined below. With respect to this topology, T_ω is uniformly expanding, and X_ω is closed in $\mathbb{N}^{\mathbb{N}}$ but in general neither compact nor locally compact.

Now assume that $\varphi : X \to \mathbb{R}$ is a measurable function such that $\varphi_\omega := \varphi|_{X_\omega} : X_\omega \to \mathbb{R}$ is locally Hölder continuous almost surely (a.s.). The Ruelle operator associated to φ is then defined by, for a.e. ω and $x \in X_{\theta\omega}$,

$$\mathscr{L}_\omega(f)(x) := \sum_{y \in X_\omega : T_\omega(y) = x} e^{\varphi_\omega(y)} f(y),$$

where $f : X_\omega \to \mathbb{R}$ is in a suitable function space such that $\mathscr{L}_\omega$ is well-defined. In [**18**], the notion of *(random) big images and preimages* was introduced in order to have a sufficient condition for a random Ruelle's theorem at hand. This condition, inspired by the one of Sarig in [**16**], requires that there exist subsets Ω_{bi} and Ω_{bp} of positive measure in the base Ω and a finite subset $\mathscr{I}$ of $\mathbb{N}$ such that:

(1) for $\theta\omega \in \Omega_{\mathrm{bi}}$ and $u \in \mathscr{W}_\omega^1$, there exists $b \in \mathscr{I}$ such that ub is ω-admissible;
(2) for $\theta\omega \in \Omega_{\mathrm{bp}}$ and $u \in \mathscr{W}_{\theta\omega}^1$, there exists $b \in \mathscr{I}$ such that bu is ω-admissible.

The basic example for a system with this property is a random topological Markov chain such that all entries of A_ω are equal to one for ω in a subset Ω^* of Ω of positive measure. In this situation, one might think of (X, T) as a random topological Markov chain that is a full shift with positive probability.

The main result of [**18**, Theorem 4.7] is the following version of Ruelle's theorem under the assumptions of local Hölder continuity of φ_ω and big images and preimages: there exist a random variable $\{\lambda_\omega\}$, a random function $\{h_\omega\}$ and a random measure $\{\mu_\omega\}$, such that for almost every $\omega \in \Omega$, $\mathscr{L}_\omega(h_\omega) = \lambda_\omega h_{\theta\omega}$, $(\mathscr{L}_\omega)^*(\mu_{\theta\omega}) = \lambda_\omega \mu_\omega$ and, for all $f_\omega \in L^1(\mu_\omega)$,

$$\lim_{n\to\infty} \left\| (\lambda_\omega \cdots \lambda_{\theta^{n-1}\omega})^{-1} \mathscr{L}_\omega^n(f^\omega) - h_{\theta^n\omega} \int f_\omega \, d\mu_\omega \right\|_{L^1(\mu_{\theta^n\omega})} = 0, \tag{1}$$

where $\mathscr{L}_\omega^n := \mathscr{L}_{\theta^{n-1}\omega} \cdots \mathscr{L}_\omega$. However, the rate of convergence remained open. The main results in here, Theorems 4.1 and 4.2, give answer to this in two ways. In Theorem 4.1, the following generalization of the results on exponential decay by Bogenschütz and

https://doi.org/10.1017/etds.2015.61 Published online by Cambridge University Press

Gundlach [**2**] and Kifer [**13**] is obtained. There exists $s \in (0, 1)$ such that for a fibrewise Lipschitz continuous function $f = \{f_\omega\}$, $n \in \mathbb{N}$, and a.e. $\omega \in \Omega$,

$$\left\| (\lambda_\omega \cdots \lambda_{\theta^{n-1}\omega})^{-1} \mathscr{L}_\omega^n(f)/h_{\theta^n(\omega)} - \int f \, d\mu_\omega \right\|_L^{\theta^n\omega} \leq \|f\|_L^\omega \cdot O_\omega(s^n), \tag{2}$$

where $\|f\|_L^\omega$ refers to the Lipschitz norm of the restriction $f|_{X_\omega}$. Note that (2) gives rise to an interpretation as a spectral gap: the operator might be written as a sum of a projection to a one-dimensional space and an operator that decays eventually exponentially fast, even though the waiting time to decay depends on the fibre X_ω. Neither the proofs in [**2**] and [**13**] nor the method used in here provide control of the random variables associated with $O_\omega(s^n)$. However, as shown in Theorem 4.2, this is possible in case that $\mathscr{L}_\omega(1) = 1$ a.s. In there, the constants associated with exponential decay along returns to Ω_{bp} are explicitly given. For illustration, these results are applied to random matrices, where it is possible to specify the associated constants under additional assumptions. Furthermore, the results are applied to decay of correlations and are used to show that the ψ-mixing coefficients of the associated process have exponential decay. As a further application, it is shown that the random measure given by $h_\omega \, d\mu_\omega$ is an equilibrium state.

The article's structure is as follows. Section 2 contains the main lemma, Lemma 2.1, which is proven in the context of non-stationary shift spaces as introduced in [**1**, **9**] in order to reveal the pathwise nature of the lemma. Namely, for a sequence of operators $(\mathscr{L}_k)$ such that each $\mathscr{L}_k$ maps Hölder functions defined on a Polish space Σ_k to Hölder functions on Σ_{k+1}, it is shown that the dual operators eventually contract the Vasershtein distance. In particular, in a strict sense, the lemma is not related to spectral theory since even the spectra of the operators are not defined. In this situation, the technique of Birkhoff cones was applied by Fisher in [**9**] in order to characterize unique ergodicity of adic transformations through a Perron–Frobenius theorem for products of transition matrices given by a non-stationary shift space. However, this technique seems to be unavailable due to the lack of compactness of Σ_k. In here, the proof relies on an argument by Hairer and Mattingly in [**11**] and is based on the Monge–Kantorovich duality, optimal transport and the construction of the coupling in step (2) of the proof of Lemma 2.1. Even though coupling methods are known in the theory of dynamical systems and were used e.g. by Kuzmin in order to show subexponential decay for the Gauss map, and by Bressaud *et al* in [**3**] to obtain decay of correlations, and the Vasershtein metric was used by Galatolo and Pacifico in [**10**] to obtain exponential decay of a return map of a Lorenz-like flow, it seems that the combination through ideas from optimal transport were not known in this context. As an application of this main lemma, it is shown in Corollary 2.1 under very mild conditions on the non-stationary shift space that invariant measures are unique and that the Ruelle operators converge.

In §3, random topological Markov chains with the big images and big preimages (b.i.p.) property are introduced. Moreover, the main lemma is adapted and applied to random topological Markov chains: the ergodicity of θ allows us to control the parameters in Lemma 2.1 through successive returns to certain subsets of Ω_{bp}. In particular, if $\mathscr{L}_\omega(1) = 1$ a.s., then this gives rise to a contraction of the Vasershtein metric in Corollaries 3.1 and 3.2. Section 4 contains two versions of Ruelle's theorem, Theorems 4.1

https://doi.org/10.1017/etds.2015.61 Published online by Cambridge University Press

and 4.2, for random topological Markov chains and their applications to random matrices, equilibrium states and asymptotic independence. For example, it is shown that the so-called ψ-mixing coefficients decay exponentially fast and that there is an exponential decay of correlations of Hölder functions against functions in L^1. Finally, in §5, Theorems 4.1 and 4.2 are proven by combining the results of §3 with the existence of $\{\lambda_\omega\}$ and $\{\mu_\omega\}$ given by [**18**, Theorem 4.7].

2. *Non-stationary shift spaces, Vasershtein distances and the main lemma*

We now introduce the notion of non-stationary shift spaces and prove the main lemma in this context in order to reveal its pathwise character. In here, the notation is adapted to that for random topological Markov chains and therefore differs from the definition in [**1**, **9**] essentially by not explicitly using Bratteli diagrams. So assume that for each $n \in \mathbb{N}_0 := \mathbb{N} \cup \{0\}$ there is a subset $\mathscr{W}_n^1 \subset \mathbb{N}$ and a matrix $A_n = (\alpha_{ij}^n,\ i \in \mathscr{W}_n^1,\ j \in \mathscr{W}_{n+1}^1)$, with entries $\alpha_{ij}^n \in \{0, 1\}$ such that $\sum_{j \in \mathscr{W}_{n+1}^1} \alpha_{ij}^n > 0$ and $\sum_{j \in \mathscr{W}_{n-1}^1} \alpha_{ji}^{n-1} > 0$ for all $i \in \mathscr{W}_n^1$ with $n \in \mathbb{N}_0$ and $n \in \mathbb{N}$, respectively. This gives rise to a sequence of shift spaces

$$\Sigma_n := \{(x_0, x_1, \ldots) : \alpha_{x_i x_{i+1}}^{i+n} = 1\ \forall i = 0, 1, \ldots\}$$

and an action $T : \Sigma_n \to \Sigma_{n+1}$, for $n \in \mathbb{N}_0$, of the shift map T defined by $T(x_0, x_1, x_2, \ldots) := (x_1, x_2, \ldots)$. In this situation, we will refer to $((\Sigma_n), T)$ as a *non-stationary shift*. If $\Sigma_n = \Sigma_m$ for all $n, m \in \mathbb{N}_0$, then we will refer to (Σ_0, T) as a *stationary shift*.

In order to define a system of neighbourhoods in analogy to stationary shifts, a finite word $a = (x_0, x_1, \ldots, x_{k-1}) \in \mathbb{N}^k$ of length k will be called *n-admissible* if $x_i \in \mathscr{W}_{n+i}^1$ and $\alpha_{x_i x_{i+1}}^{n+i} = 1$, for $i = 0, \ldots, k-1$. We will refer to $\mathscr{W}_n^k$ as the set of n-admissible words of length k and to, for $a = (a_0, a_1, \ldots, a_{k-1}) \in \mathscr{W}_n^k$,

$$[a]_n = [a_0, a_1, \ldots, a_{k-1}]_n := \{x \in \Sigma_n : x_i = a_i,\ i = 0, 1, \ldots, k-1\}$$

as a *cylinder set*. As in case of a stationary shift, it follows that Σ_n is a Polish space with respect to the topology generated by these cylinder sets and that, for $r \in (0, 1)$, the metric d_r defined by

$$d_r((x_0, x_1, x_2, \ldots), (y_0, y_1, y_2, \ldots)) := r^{\min\{i : x_i \neq y_i\}}$$

is compatible with this topology. Also note that, for $a = (a_0, a_1, \ldots, a_{k-1}) \in \mathscr{W}_n^k$, the restriction $T^k|_{[a]_n} : [a]_n \to T^k([a]_n) \subset \Sigma_{k+n}$ is a homeomorphism. The inverse of this map will be denoted by τ_a, and it can be seen easily that

$$\tau_a : T^k([a]_n) \to [a]_n,\ (x_0 x_1 \ldots) \mapsto (a_0 a_1 \ldots a_{k-1} x_0 x_1 \ldots).$$

Topological transitivity, mixing and the b.i.p. property. Incontrast to stationary shift spaces, the notions of topological transitivity etc. require an additional condition which guarantees that each element a $\in \mathscr{W}_n^1$ occurs infinitely often. That is, we refer to $((\Sigma_n), T)$ as *topologically transitive* if for all $a \in \mathscr{W}_{k_1}^1$, $b \in \mathscr{W}_{k_2}^1$ for some $k_1, k_2 \in \mathbb{N}_0$, there exists a sequence $(n_l : l \in \mathbb{N})$ with $n_l \nearrow \infty$ such that $b \in \mathscr{W}_{k+n_l}^1$ and

$$[a]_{k_1} \cap T^{-n_l}([b]_{k+n_l}) \neq \emptyset \quad \text{for all } l \in \mathbb{N}.$$

https://doi.org/10.1017/etds.2015.61 Published online by Cambridge University Press

Furthermore, we will refer to $((\Sigma_n), T)$ as *topologically mixing* if $((\Sigma_n), T)$ is topologically transitive and if, for all $a \in \mathscr{W}_{k_1}^1$, $b \in \mathscr{W}_{k_2}^1$ for some $k_1, k_2 \in \mathbb{N}_0$, there exists N_{ab} such that

$$[a]_{k_1} \cap T^{-n}([b]_{k+n}) \neq \emptyset \quad \text{for all } n \geq N_{ab} \text{ s.t. } b \in \mathscr{W}_{k+n}^1.$$

Finally, we will say that $((\Sigma_n), T)$ has *big preimages (b.p.)* if $((\Sigma_n), T)$ is topologically mixing and there exists a finite subset $\mathscr{I}$ of $\mathbb{N}$ and an infinite subset $\mathscr{K} \subset \mathbb{N}$ such that for all $n \in \mathscr{K}$ and for each $a \in \mathscr{W}_n^1$, there exists $b \in \mathscr{I} \cap \mathscr{W}_{n-1}^1$ with $ba \in \mathscr{W}_{n-1}^2$.

Hölder continuity and the Ruelle operator. For $f : \Sigma_n \to \mathbb{R}$ and $k \in \mathbb{N}$,

$$V_k(f) = \sup\{|f(x) - f(y)| : x, y \in [a]_n, a \in \mathscr{W}_n^k\}$$

is called the kth variation of f. We refer to f as a *locally (r, m)-Hölder continuous function with Hölder parameter $r \in (0, 1)$ and Hölder index $m \in \mathbb{N}$* if there exists $\kappa \geq 0$ such that $V_k(f) \leq \kappa r^k$ for all $k \geq m$. If κ is minimal, then κ will be referred to as the *Hölder constant of index m* of f. Furthermore, if f is locally (r, m)-Hölder continuous and $\|f\|_\infty < \infty$, then we will refer to f as an *r-Hölder continuous function*. Note that $\|f\|_\infty < \infty$ and local Hölder continuity imply local Hölder continuity of index 1. If we consider Σ_n equipped with the metric d_r, then we will also refer to an r-Hölder continuous function as a *Lipschitz continuous function*. In order to recover the usual definition of local Hölder continuity, it suffices to fix a metric d_{r^*} on Σ_n: the above Hölder condition is equivalent to $|f(x) - f(y)| \leq \kappa d_{r^*}(x, y)^s$, for all x, y with $d(x, y) \leq (r^*)^k$ and $s = \log r / \log r^*$.

Now assume that $(\varphi_n : \Sigma_n \to \mathbb{R})$ is a sequence of locally (r, l)-Hölder continuous functions with Hölder constant $\kappa_n > 0$. As in the stationary case, the estimate

$$S_n(\varphi_k) := \sum_{i=0}^{n-1} \varphi_{i+k} \circ T^i$$

for Birkhoff sums holds. Namely, for $n < m - l$, and $x, y \in [a]_k$ with $a \in \mathscr{W}_k^m$, we have that

$$|S_n\varphi_k(x) - S_n\varphi_k(y)| \leq \sum_{i=0}^{n-1} \kappa_{k+i} r^{m-i} \leq r^{m-n} \log B_{k+n}$$

$$\text{for } B_{k+n} := \exp \sum_{i=1}^{k+n} \kappa_{k+n-i} r^i. \tag{3}$$

This estimate allows us to identify the relevant function spaces for the following sequence of operators. Assume that $(\varphi_k : k \in \mathbb{N}_0)$ has Hölder index 2. The kth Ruelle operator with respect to the potential φ_k is defined by, for $x \in \Sigma_{k+1}$ and $f : \Sigma_k \to \mathbb{R}$ in a suitable space,

$$\mathscr{L}_k(f)(x) = \sum_{Ty=x} e^{\varphi_k(y)} f(y).$$

For $k \in \mathbb{N}$, set $\mathscr{L}_k^n := \mathscr{L}_{k+n-1} \circ \cdots \circ \mathscr{L}_{k+1} \circ \mathscr{L}_k$ and note that $\mathscr{L}_n^k$ maps functions with domain Σ_n to functions with domain Σ_{n+k} and that

$$\mathscr{L}_k^n(f)(x) = \sum_{T^n(y)=x} e^{S_n\varphi_k(y)} f(y) = \sum_{v \in \mathscr{W}_k^n} e^{S_n\varphi_k(\tau_v(x))} f(\tau_v(x)).$$

https://doi.org/10.1017/etds.2015.61 Published online by Cambridge University Press

PROPOSITION 2.1. *If $\mathscr{L}_k(1) = 1$ and (φ_k) is locally $(r, 2)$-Hölder for all $k \in \mathbb{N}_0$, then $\mathscr{L}_k^n$ maps r-Hölder continuous functions defined on Σ_k to r-Hölder continuous functions defined on Σ_{k+n}. In particular, for a Hölder continuous function $f : \Sigma_k \to \mathbb{R}$, we have $\|\mathscr{L}_k^n(f)\|_\infty \le \|f\|_\infty$ and the Hölder constant of index 1 of $\mathscr{L}_k^n(f)$ is smaller than or equal to $\kappa_f r^n + \|f\|_\infty(B_{k+n} - 1)$, with κ_f referring to the Hölder constant of index $n + 1$ of f.*

Proof. The estimate (3) gives rise to, for $x, y \in \Sigma_{n+k}$ with $d_r(x, y) \le r$,

$$\begin{aligned}
&|\mathscr{L}_k^n(f)(x) - \mathscr{L}_k^n(f)(y)| \\
&\quad \le \sum_{v \in \mathscr{W}_k^n} e^{S_n\varphi_k(\tau_v(x))}|f(\tau_v(x)) - f(\tau_v(y))| \\
&\qquad + \sum_{v \in \mathscr{W}_k^n} e^{S_n\varphi_k(\tau_v(x))}|f(\tau_v(y))(1 - e^{S_n\varphi_k(\tau_v(y)) - S_n\varphi_k(\tau_v(x))})| \\
&\quad \le \sup_{v \in \mathscr{W}_k^n}\left(\kappa_f d_r(\tau_v(x), \tau_v(y)) + \frac{\|f\|_\infty(B_{k+n} - 1)}{\log B_{k+n}}|S_n\varphi_k(\tau_v(y)) - S_n\varphi_k(\tau_v(x))|\right) \\
&\quad \le (\kappa_f r^n + \|f\|_\infty(B_{k+n} - 1))\, d_r(x, y),
\end{aligned}$$

where the second step relies on $|1 - \exp(x)| \le \exp(|x|) - 1$ and the monotonicity of $(\exp(x) - 1)/x$ for $x \ge 0$. Futhermore, $\|\mathscr{L}_k^n(f)\|_\infty \le \|f\|_\infty$ follows from the fact that $\mathscr{L}_k^n$ maps positive functions to positive functions and $\mathscr{L}_k^n(1) = 1$. □

Before analysing the action of the dual of $\mathscr{L}_k^n$ on Borel probability measures, we now fix the systems under consideration. That is, we say that $((\Sigma_n), T, (\varphi_n))$ satisfies *property (H*)* if $((\Sigma_n), T)$ satisfies the b.p. property, $\mathscr{L}_n(1) = 1$ for all $n \in \mathbb{N}_0$, and the potentials φ_n are locally $(r, 2)$-Hölder continuous for all $n \in \mathbb{N}_0 \backslash \mathscr{K}$ and locally $(r, 1)$-Hölder continuous for all $n \in \mathscr{K}$.

2.1. *Vasershtein distances and the coupling construction.* Assume that Y is a Polish space and that μ, ν are Borel probability measures on Y. We then refer to

$$\Pi(\mu, \nu) := \{m \in \mathscr{P}(Y^2) : \pi_1^*(m) = \mu, \pi_2^*(m) = \nu\}$$

as the couplings of μ and ν, where $\mathscr{P}(Y^2)$ is the set of Borel probability measures on Y^2 and π_i are the canonical projections. Now assume that d is a metric compatible with the topology of Y. The Vasershtein distance of μ and ν is then defined by

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

As is well-known (see, e.g., [**21**]), the Vasershtein distance is compatible with weak convergence and, by the Monge–Kantorovich duality,

$$W(\mu, \nu) = \sup\left\{\int f\, d\mu - \int f\, d\nu : D(f) \le 1\right\}, \tag{4}$$

where $D(f) := \sup\{|f(x) - f(y)|/d(x, y) : x, y \in Y\}$ refers to the *Lipschitz constant* of f. Note that $D(f)$ is obtained, in contrast to the Hölder constant defined above, by taking the supremum over all pairs (x, y).

https://doi.org/10.1017/etds.2015.61 Published online by Cambridge University Press

The following proposition is the principal result of this section and is obtained by adapting the asymptotic coupling method in [**11**, §2.1] to the non-stationary setting. Namely, as a corollary of Lemma 2.1, the operator defined by

$$\int f \, d\mathscr{L}_k^{n*}(\mu) := \int \mathscr{L}_k^n(f) \, d\mu$$

maps Borel probability measures on Σ_{k+n} to Borel probability measures on Σ_k. By applying an asymptotic coupling method and the Monge–Kantorovich duality, it is shown below that $\mathscr{L}_k^{n*}$ acts as a contraction with respect to the Vasershtein distance. In order to shed light on the approach via the coupling construction, the involved constants are subject to a careful analysis in order to obtain effective bounds with respect to this method.

The associated constants are defined as follows. Let $\beta \in (0, 1)$, let r be the Hölder parameter of $\{\varphi_k : k \in \mathbb{N}_0\}$, and let B_k be as in (3). For $k \in \mathbb{N}_0$, define

$$\alpha_k := B_k/\beta, \quad \overline{d}_k(x, y) := \min\{1, \alpha_k d_r(x, y)\},$$

giving rise to the Vasershtein distances W_k and Lipschitz constants $\overline{D}_k$ with respect to the metric $\overline{d}_k$ on Σ_k. Also choose $n_k \in \mathbb{N}$ with

$$n_k \geq \lfloor -\log \alpha_k / \log r \rfloor + 1,$$

where $\lfloor t \rfloor$ refers to the biggest integer smaller than or equal to $t \in \mathbb{R}$. Furthermore, for each $k \in \mathbb{N}_0$, choose an element $\mathbf{o}_k$ in $\mathscr{W}_k^1$ and define

$$m_k := \min(\{n \geq 1 : k + n \in \mathscr{K}, (A_k A_{k+1} \cdots A_{k+n-2})_{\mathbf{o}_k j} \geq 1 \; \forall j \in \mathscr{J} \cap \mathscr{W}_{k+n-1}^1\}),$$

which is finite by the topological mixing property. Now assume that $x \in \Sigma_{k+m_k}$. By the b.p. property, there is $j(x) \in \mathscr{J}$ such that $x \in T([j(x)]_{k+m_k-1})$. Moreover, since $\mathscr{J}$ is a finite set, there exists a finite subset $\mathscr{U}_k \subset \mathscr{W}_k^{m_k-1}$ such that for each $j \in \mathscr{J} \cap \mathscr{W}_{k+m_k-1}^1$, there exists an element $u = (\mathbf{o}_k u_1 \ldots j) \in \mathscr{U}_k$. Hence, for each $x \in \Sigma_{k+m_k}$, there exists $u(x) \in \mathscr{U}_k$ with $x \in T^{m_k}([u(x)]_k)$. In particular, it follows that

$$C_k := \inf_{x \in \Sigma_{k+m_k}} e^{S_{m_k}\varphi_k(\tau_{u(x)}x)} \geq \inf_{u \in \mathscr{U}_k} \inf_{y \in [u]_k} e^{S_{m_k}\varphi_k(y)} > 0. \tag{5}$$

And finally, for $k \in \mathbb{N}_0$, set

$$t_k := \max\left\{\beta, \, 1 - \frac{(1 - r^{n_k}\alpha_k)C_{k+n_k}}{B_{k+n_k}}\right\} \in (0, 1). \tag{6}$$

LEMMA 2.1. *Assume that* $((\Sigma_n), T, (\varphi_n))$ *satisfies* (H^*) *and that* r *is the Hölder parameter of* φ_k. *Furthermore, assume that for each* $k \in \mathbb{N}_0$, $f_k : \Sigma_k \to \mathbb{R}$ *is* r*-Hölder continuous and that* μ_k *and* ν_k *are Borel probability measures on* Σ_k. *With* $l_k := n_k + m_{k+n}$, *we then have that:*

(1) $\overline{D}_{k+n_k}(\mathscr{L}_k^{n_k}(f_k)) \leq \overline{D}_k(f_k)$;

(2) $\overline{D}_{k+l_k}(\mathscr{L}_k^{l_k}(f_k)) \leq t_k \cdot \overline{D}_k(f_k)$;

(3) $W_k((\mathscr{L}_k^{n_k})^*(\mu_{k+n_k}), (\mathscr{L}^{n_k})^*(\nu_{k+n_k})) \leq W_{k+n_k}(\mu_{k+n_k}, \nu_{k+n_k})$;

(4) $W_k((\mathscr{L}_k^{l_k})^*(\mu_{k+l_k}), (\mathscr{L}_k^{l_k})^*(\nu_{k+l_k})) \leq t_k \cdot W_{k+l_k}(\mu_{k+l_k}, \nu_{k+l_k})$.

Proof. The proof of assertions (2) and (4) consists of three main steps. We derive first a local contraction of $(\mathscr{L}_k^n)^*$, then a concentration of $(\mathscr{L}_k^n)^*$-images of Dirac-measures close to the diagonal, and finally extend the results to arbitrary measures. The remaining, easier assertions then are proved using a simplified version of the second step.

https://doi.org/10.1017/etds.2015.61 Published online by Cambridge University Press

Step 1: Local contraction of $\mathscr{L}$. For a Lipschitz continuous function f_k, we clearly have that

$$|f_k(x) - f_k(y)|/\overline{d}_k(x, y) \le \begin{cases} D(f_k)/\alpha_k : d_r(x, y) < 1/\alpha_k \\ D(f_k) : d_r(x, y) \ge 1/\alpha_k. \end{cases}$$

In particular, $\overline{D}_k(f_k) \le D(f_k) \le \alpha_k \overline{D}_k(f_k)$, which proves that f_k is Lipschitz continuous on Σ_k with respect to the metric $\overline{d}_k$. We now aiming for an estimate of $|\mathscr{L}_\omega^n(f)(x) - \mathscr{L}_\omega^n(f)(y)|$. Using $\mathscr{L}_k(c) = c$, we may assume without loss of generality that $\inf\{f_k(x) : x \in \Sigma_k\} = 0$. Hence, we may assume that $\|f_k\|_\infty \le \overline{D}_k(f_k)$. Combining $r^{n_\omega}\alpha_\omega < 1$ with Proposition 2.1 then gives that, for $x, y \in \Sigma_{k+n_k}$ with $d_r(x, y) \le r$,

$$\begin{aligned} \frac{|\mathscr{L}_k^{n_k}(f_k)(x) - \mathscr{L}_k^{n_k}(f_k)(y)|}{\alpha_{k+n_k} d_r(x, y)} &\le \frac{D(f_k) r^n d_r(x, y)}{\alpha_{k+n_k} d_r(x, y)} + \frac{\|f_k\|_\infty (B_{k+n_k} - 1)}{\alpha_{k+n_k}} \\ &\le \frac{\overline{D}_k(f_k)(\alpha_k r^{n_k} + B_{k+n_k} - 1)}{\alpha_{k+n_k}} \\ &\le \frac{(\overline{D}_k(f_k) B_{k+n_k})}{\alpha_{k+n_k}}. \end{aligned} \tag{7}$$

By the choice of α_{k+n}, the right-hand side of (7) is smaller than or equal to $\beta \cdot \overline{D}_k(f_k)$.

Step 2: Concentration near the diagonal. We now construct a coupling $Q_{x,y} \in \Pi((\mathscr{L}_k^{l_k})^*(\delta_x), (\mathscr{L}_k^{l_k})^*(\delta_y))$ with a uniform lower bound on a neighbourhood of the diagonal by considering inverse branches starting in $\mathscr{K}$ and passing through the cylinder $[\mathbf{o}]_{k+n_k} := [\mathbf{o}_{k+n_k}]_{k+n_k}$. Namely, the coupling is defined through the following decomposition. For $x \in \Sigma_{k+l_k}$ and $v = (v_1 v_2) \in \mathscr{W}_k^{l_k}$ with $v_1 \in \mathscr{W}_k^{n_k}$ and $v_2 \in \mathscr{W}_{k+n_k}^{m_{k+n_k}}$, define

$$\begin{aligned} \phi_v^{(1)}(x) &:= \mathbf{1}_{[u(x)]_{k+n_k}} \circ T^{n_k}(x) \cdot \inf(\{e^{S_{l_k}(\varphi_k)\circ\tau_{v_1 u(y)}(y)} : y \in \Sigma_{k+l_k}\}), \\ \phi_v^{(2)}(x) &:= e^{S_{l_k}(\varphi_k)\circ\tau_v(x)} - \phi_v^{(1)}(x), \end{aligned}$$

Observe that the indicator in the definition of $\phi_v^{(1)}(x)$ is equal to one if and only if $v_2 = u(x)$ and that $\phi_{v_1 u(x)}^{(1)}(x) = \phi_{v_1 u(y)}^{(1)}(y)$ for all $x, y \in \Sigma_{k+l_k}$. In order to obtain a bound on $\phi_v^{(1)}(x)$ for $v_2 = u(x)$, note that it follows from (5) that

$$\phi_v^{(1)}(x) \ge C_{k+n_k} \inf_{y \in [\mathbf{o}]_{k+n_k}} e^{S_{n_k}(\varphi_k)\circ\tau_{v_1}(y)} \ge \frac{C_{k+n_k}}{B_{k+n_k}} \sup_{y \in [\mathbf{o}]_{k+n_k}} e^{S_{n_k}(\varphi_k)\circ\tau_{v_1}(y)}. \tag{8}$$

This gives rise to a probability measure $Q_{x,y}$ on $\Sigma_k \times \Sigma_k$ by

$$\begin{aligned} Q_{x,y} := &\sum_{v_1 \in \mathscr{W}_k^{n_k}} \phi_{v_1 u(x)}^{(1)}(x) \delta_{(\tau_{v_1 u(x)}(x), \tau_{v_1 u(y)}(y))} \\ &+ \frac{1}{1 - \sum_{v_1 \in \mathscr{W}_k^{n_k}} \phi_{v_1 u(x)}^{(1)}(x)} \sum_{v, w \in \mathscr{W}_k^{l_k}} \phi_v^{(2)}(x) \phi_w^{(2)}(y) \delta_{(\tau_v(x), \tau_w(y))}. \end{aligned}$$

As can be verified easily, $Q_{x,y}$ is a coupling of $(\mathscr{L}_k^{l_k})^*(\delta_x)$ and $(\mathscr{L}_k^{l_k})^*(\delta_x)$. By construction, we have that $\overline{d}_k(\tau_v(x), \tau_v(y)) \le r^{n_k}\alpha_k < 1$ for all $v \in \mathscr{W}_k^{n_k}$ and $x, y \in$

https://doi.org/10.1017/etds.2015.61 Published online by Cambridge University Press

$[\mathbf{o}]_{k+n_k}$. Hence, by (8) and $\mathscr{L}_k^{n_k}(1) = 1$,

$$Q_{x,y}(\{(z, z') : \overline{d}_k(z, z') \le r^{n_k}\alpha_k\}) \ge \frac{C_{k+n_k}}{B_{k+n_k}}.$$

We will now employ this estimate in order to obtain uniform contraction of $\overline{D}(f)$. In order to do so, let $\Delta_k := \{(x', y') \in \Sigma_k \times \Sigma_k : \overline{d}_k(x', y') \le r^{n_k}\alpha_k\}$. For $x, y \in \Sigma_{k+l_k}$, we then have

$$\begin{aligned} W_k((\mathscr{L}_k^{l_k})^*(\delta_x), (\mathscr{L}_k^{l_k})^*(\delta_y)) &\le \int \overline{d}_k(x', y')\, dQ_{x,y}(x', y') \\ &\le r^{n_k}\alpha_k Q_{x,y}(\Delta_k) + 1 - Q_{x,y}(\Delta_k) \\ &= 1 - (1 - r^{n_k}\alpha_k)Q_{x,y}(\Delta_k) \\ &\le 1 - \frac{(1 - r^{n_k}\alpha_k)C_{k+n_k}}{B_{k+n_k}} =: s_k < 1. \end{aligned}$$

Hence, by the Monge–Kantorovich duality,

$$|\mathscr{L}_k^{l_k}(f_k)(x) - \mathscr{L}_k^{l_k}(f_k)(y)| = \left|\int f_k\, d(\mathscr{L}_k^{l_k})^*(\delta_x) - \int f_k\, d(\mathscr{L}_k^{l_k})^*(\delta_y)\right| \le \overline{D}_k(f_k) \cdot s_k.$$

Step 3: Contraction of the Lipschitz constant and extension to arbitrary measures. By the above,

$$\frac{|\mathscr{L}_k^{l_k}(f)(x) - \mathscr{L}_k^{l_k}(f)(y)|}{\overline{d}_{k+l_k}(x, y)} \le \begin{cases} s_k \cdot \overline{D}_k(f) : \overline{d}_{k+l_k}(x, y) = 1 \\ \beta \cdot \overline{D}_k(f) : \overline{d}_{k+l_k}(x, y) < 1. \end{cases}$$

Hence, for $t_k = \max\{\beta, s_k\}$, we have $\overline{D}_{k+l_k}(\mathscr{L}_k^{l_k}(f)) \le t_k \cdot \overline{D}_k(f)$, which proves assertion (2) of the lemma. By the Monge–Kantorovich duality, this implies that

$$W_k((\mathscr{L}_k^{l_k})^*(\delta_x), (\mathscr{L}_k^{l_k})^*(\delta_y)) \le t_k \overline{d}_{k+l_k}(x, y).$$

Since μ_{k+l_k} and ν_{k+l_k} are probability measures on Σ_{k+l_k}, there exists a coupling Q in $\Pi(\mu_{k+l_k}, \nu_{k+l_k})$, referred to as optimal transport, such that $W_{k+l_k}(\mu_{k+l_k}, \nu_{k+l_k}) = \int \overline{d}_{k+l_k}(x, y)\, dQ(x, y)$. Moreover, let $P_{x,y} \in \Pi((\mathscr{L}_k^{l_k})^*(\delta_x), (\mathscr{L}_k^{l_k})^*(\delta_y))$ refer to an optimal transport of $(\mathscr{L}_k^{l_k})^*(\delta_x)$ and $(\mathscr{L}_k^{l_k})^*(\delta_y)$. Using a construction similar to the one in Step 2, it is then possible to show the following. For $\varepsilon > 0$ there exists $\delta > 0$ and a coupling $P^{\varepsilon}_{x',y'} \in \Pi((\mathscr{L}_k^{l_k})^*(\delta_{x'}), (\mathscr{L}_k^{l_k})^*(\delta_{y'}))$ constructed from $P_{x,y}$ in a continuous way, such that

$$\left|\int \overline{d}_k\, dP_{x,y} - \int \overline{d}_k\, dP^{\varepsilon}_{x',y'}\right| \le \varepsilon \quad \text{for all } x', y' : d_r(x, x'), d_r(y, y') < \delta.$$

In particular, $(x, y) \to W_k((\mathscr{L}_k^{l_k})^*(\delta_x), (\mathscr{L}_k^{l_k})^*(\delta_y))$ is continuous and there exists a locally continuous family $\{P^{\varepsilon}_{x,y} : x, y \in \Sigma_{k+l_k}\}$ that approximates the Vasershtein distance up to ε. For Q^{ε} defined by $dQ^{\varepsilon}(x, y) := P^{\varepsilon}_{x,y}\, dQ(x, y)$, it is then easy to see that

https://doi.org/10.1017/etds.2015.61 Published online by Cambridge University Press

$Q^\varepsilon \in \Pi((\mathscr{L}_k^{l_k})^*(\mu_{k+l_k}), (\mathscr{L}_k^{l_k})^*(\nu_{k+l_k}))$. Hence,

$$\begin{aligned} W_k((\mathscr{L}_k^{l_k})^*(\mu_{k+l_k}), (\mathscr{L}_k^{l_k})^*(\nu_{k+l_k})) &\le \int \left(\int \overline{d}_k \, dP^\varepsilon_{x,y} \right) dQ(x, y) \\ &\le \int W_k((\mathscr{L}_k^{l_k})^*(\delta_x), (\mathscr{L}_k^{l_k})^*(\delta_y)) \, dQ(x, y) + \varepsilon \\ &\le t_k \int \overline{d}_{k+l_k}(x, y) \, dQ(x, y) + \varepsilon \\ &= t_k W_{k+l_k}(\mu_{k+l_k}, \nu_{k+l_k}) + \varepsilon. \end{aligned}$$

Since $\varepsilon > 0$ is arbitrary, assertion (4) follows.

Step 4: Proof of assertions (1) and (3) of Lemma 2.1. The proof uses a simplified version of the arguments above. As in Step 2, we have

$$W_k((\mathscr{L}_k^{n_k})^*(\delta_x), (\mathscr{L}_k^{n_k})^*(\delta_y)) \le \int \overline{d}_k(x', y') \, dQ_{x,y}(x', y') \le 1.$$

The assertions then follow by the same argument as in Step 3. □

As a consequence of the above lemma, we obtain the following result on the convergence of $\mathscr{L}_k^n$ and the unicity of the invariant measures for non-stationary shift spaces with property (H*). In here, we refer to a sequence of probability measures (μ_k) as an *invariant sequence* if $\mathscr{L}_k^*(\mu_{k+1}) = \mu_k$ for all $k \in \mathbb{N}_0$.

COROLLARY 2.1. *Assume that* $((\Sigma_n), T, (\varphi_n))$ *satisfies (H*), that there exists an invariant sequence* (μ_k) *of probability measures, and that* n_k *might be chosen such that* t_k *in (6) is uniformly bounded away from* 1. *Then the sequence* (μ_k) *is unique and moreover, for a Lipschitz function* $f : \Sigma_k \to \mathbb{R}$, *we have that*

$$\left\| \mathscr{L}_k^n f - \int f \, d\mu_k \right\|_\infty \xrightarrow{n\to\infty} 0. \tag{9}$$

Proof. By hypothesis, $t := \sup_n t_n < 1$. Lemma 2.1 then gives that, for $k \in \mathbb{N}_0$ and the sequence inductively defined by $p_0 = k$ and $p_{j+1} = p_j + l_{p_j}$,

$$W_k((\mathscr{L}_k^{p_j-k})^*(\mu_{p_j}), (\mathscr{L}_k^{p_j-k})^*(\nu_{p_i})) \le t^j W_{p_j}(\mu_{p_j}, \nu_{p_j}) \le t^j.$$

Hence, if (μ_k) and (ν_k) are invariant sequences of probability measures, then $W_k(\mu_k, \nu_k) = 0$. That is, the sequence (μ_k) is unique. Now assume that f is Lipschitz and that, without loss of generality, $\overline{D}_k(f) \le 1$. We then have by the Monge–Kantorovich duality and using the same argument as above that, for $x \in \Sigma_{p_j}$,

$$\begin{aligned} \left| \mathscr{L}_k^{p_j-k}(f)(x) - \int f \, d\mu_k \right| &= \left| \int f \, d(\mathscr{L}_k^{p_j-k})^*(\delta_x) - \int f \, d(\mathscr{L}_k^{p_j-k})^*(\mu_{p_j}) \right| \\ &\le W_k((\mathscr{L}_k^{p_j-k})^*(\delta_x), (\mathscr{L}_k^{p_j-k})^*(\mu_{p_j})) \le t^j. \end{aligned}$$

Hence, $\|\mathscr{L}_k^n f - \int f \, d\mu_k\|_\infty \to 0$ along a subsequence. However, since $\mathscr{L}_k(1) = 1$, we have that $\|\mathscr{L}_k(f)\|_\infty \le \|f\|_\infty$ for all k, which proves the assertion. □

https://doi.org/10.1017/etds.2015.61 Published online by Cambridge University Press

With respect to the corollary and its the proof, it is worth noting that the corollary is applicable in many situations since (n_k) might grow arbitrarily fast. However, this implies also that the speed of convergence in (9) could be arbitrarily slow. For more specific applications of the main lemma, one might consider situations with several uniform bounds. For example, if $\mathscr{K}$ has a positive density and the mixing times m_k, the Hölder constants of (φ_k) and the C_k in (5) are uniformly bounded (e.g. by a condition establishing some type of almost stationarity of (φ_k) restricted to $\mathscr{U}_k$), then the main lemma would imply exponential speed of convergence to the limit in (9). However, for random countable topological Markov chains, it is shown below that the analogues of these conditions are automatically satisfied due to the ergodicity of θ. Or, from an abstract point of view, the ergodicity of θ establishes a sufficient level of stationarity.

3. *Random countable topological Markov chains*

In this section, the details of the construction of random countable topological Markov chains and Hölder potentials are given. Note that the main difference to the definitions and arguments for non-stationary shift spaces rely in the underlying measurable structure. For the definition of a random countable topological Markov chain, assume that θ is an ergodic automorphism (i.e. ergodic, bimeasurable, invertible and probability preserving) of the probability space $(\Omega, \mathscr{F}, P)$, that $\{\mathscr{W}_\omega^1 : \omega \in \Omega\}$ is a measurable family of subsets of $\mathbb{N}$ and that, for a.e. $\omega \in \Omega$, $A_\omega = (\alpha_{ij}^\omega,\ i \in \mathscr{W}_\omega^1,\ j \in \mathscr{W}_{\theta\omega}^1)$ is a matrix with entries $\alpha_{ij}^\omega \in \{0, 1\}$ such that $\omega \mapsto A_\omega$ is measurable and $\sum_{j\in\mathscr{W}_{\theta\omega}^1} \alpha_{ij}^\omega > 0$, $\sum_{j\in\mathscr{W}_{\theta^{-1}\omega}^1} \alpha_{ji}^{\theta^{-1}\omega} > 0$ for all $i \in \mathscr{W}_\omega^1$. For the random shift space

$$X_\omega = \{(x_0, x_1, \ldots) : \alpha_{x_i x_{i+1}}^{\theta^i\omega} = 1\ \forall i = 0, 1, \ldots\},$$

the (random) shift map $T_\omega : X_\omega \to X_{\theta\omega}$ is defined by $T_\omega : (x_0, x_1, x_2, \ldots) \mapsto (x_1, x_2, \ldots)$. This gives rise to a globally defined map $T : X \to X$, with $X := \{(\omega, x) : x \in X_\omega\}$ and $T(\omega, x) = (\theta\omega, T_\omega x)$. The quintuple $(X, T, \Omega, P, \theta)$, sometimes abbreviated by (X, T), then is referred to as a *random countable topological Markov chain*.

A finite word $a = (x_0, x_1, \ldots, x_{n-1}) \in \mathbb{N}^n$ of length n is called *ω-admissible* if $x_i \in \mathscr{W}_{\theta^i\omega}^1$ and $\alpha_{x_i x_{i+1}}^{\theta^i\omega} = 1$, for $i = 0, \ldots, n-1$. Moreover, $\mathscr{W}_\omega^n$ denotes the set of ω-admissible words of length n and, for $a = (a_0, a_1, \ldots, a_{n-1}) \in \mathbb{N}^n$,

$$[a]_\omega = [a_0, a_1, \ldots, a_{n-1}]_\omega := \{x \in X_\omega : x_i = a_i,\ i = 0, 1, \ldots, n-1\}$$

is called a *cylinder set*. As can be verified easily, X_ω is a closed subset of $\mathbb{N}^{\mathbb{N}}$ with respect to the topology generated by cylinder sets, and hence X_ω is a Polish space. As in case of non-stationary shift spaces, the shift metric d_r is compatible with this topology for all $r \in (0, 1)$. As above, $T_\omega^n := T_{\theta^{n-1}\omega} \circ \cdots \circ T_{\theta\omega} \circ T_\omega$ is a homeomorphism from $[a]_\omega$ onto $T_{\theta^{n-1}\omega}([a_{n-1}]_{\theta^{n-1}\omega})$ whose inverse will be denoted by

$$\tau_a : T_\omega^n([a]_\omega) \to [a]_\omega, \quad (x_0 x_1 \ldots) \mapsto (a x_0 x_1 \ldots).$$

The set of those $\omega \in \Omega$ where the cylinder is non-empty will be denoted by Ω_a, that is,

$$\Omega_a = \{\omega : [a]_\omega \neq \emptyset\} = \{\omega : a \in \mathscr{W}_\omega^n\}.$$

The set $\mathscr{W}^n$ refers to the set of words a of length n with $P(\Omega_a) > 0$.

https://doi.org/10.1017/etds.2015.61 Published online by Cambridge University Press

For the definition of random probability measures, we will adapt the definition in [**5**] to our setting. This adaption relies on the fact that the measurable structure of $\{(x, \omega) : x \in X_\omega, \omega \in \Omega\}$ is induced by the product structure of $\mathbb{N}^{\mathbb{N}} \times \Omega$. That is, we refer to $\mu = \{\mu_\omega\}$ as a random probability measure if μ is a map

$$\mu : \mathscr{B} \times \Omega \to [0, 1], \quad (B, \omega) \mapsto \mu_\omega(B),$$

where $\mathscr{B}$ refers to the Borel σ-algebra of the countable full shift $\mathbb{N}^{\mathbb{N}}$, such that:

(1) for every $B \in \mathscr{B}$, $\omega \mapsto \mu_\omega(B)$ is measurable;

(2) for P-a.e. $\omega \in \Omega$, $B \mapsto \mu_\omega(B)$ is a probability measure with support X_ω.

The b.i.p. property. With the notion of cylinders at hand, we now give the definitions of topologically mixing and big images and preimages. We say that (X, T) is *(fibrewise) topologically mixing* if for all $a, b \in \mathscr{W}^1$, there exists an $\mathbb{N}$-valued random variable $N_{ab} = N_{ab}(\omega)$ such that, for $n \geq N_{ab}(\omega)$, $a \leq \mathscr{W}^1_\omega$ and $\theta^n\omega \in \Omega_b$, we have that $[a]_\omega \cap (T^n_\omega)^{-1}[b]_{\theta^n\omega} \neq \emptyset$.

Morever, assume that there exist $\Omega_{\mathrm{bi}} \subset \Omega$ and $\Omega_{\mathrm{bp}} \subset \Omega$ of positive measure and a finite subset $\mathscr{I}$ of $\mathbb{N}$ such that:

(1) for each $\omega \in \Omega_{\mathrm{bi}}$ and $a \in \mathscr{W}^1_{\theta^{-1}\omega}$, there exists $b \in \mathscr{I}$ with $ab \in \mathscr{W}^2_{\theta^{-1}\omega}$;

(2) for each $\omega \in \Omega_{\mathrm{bp}}$ and $a \in \mathscr{W}^1_\omega$, there exists $b \in \mathscr{I}$ with $ba \in \mathscr{W}^2_{\theta^{-1}\omega}$.

If, in addition, (X, T) is topologically mixing, then (X, T) is said to have the *(relative) big images and big preimages property* or *(relative) b.i.p. property*. Note that this definition is equivalent to one in [**18**].

Hölder and Lipschitz continuity. In order to define fibrewise Hölder continuity, the nth variation of a function $f : X \to \mathbb{R}$, $(\omega, x) \mapsto f_\omega(x)$ is defined by

$$V^\omega_n(f) = \sup\{|f_\omega(x) - f_\omega(y)| : x_i = y_i, i = 0, 1, \ldots, n-1\}.$$

We refer to f as a *(fibrewise) locally (r, k)-Hölder continuous function with parameter $r \in (0, 1)$ and index $k \in \mathbb{N} \cup \{0\}$* if there exists a random variable $\kappa = \kappa(\omega) \geq 0$ such that $\int \log \kappa \, dP < \infty$ and $V^\omega_n(f) \leq \kappa(\omega) r^n$ for all $n \geq k$. Furthermore, we will refer to a locally Hölder continuous function with $\|f_\omega\|_\infty < \infty$ a.s. as a *(fibrewise) Hölder continuous function*. Note that a Hölder continuous function is Lipschitz with respect to d_r. Therefore, we refer to

$$D_\omega(f) := \sup\{(f(x) - f(y))/d_r(x, y) : x, y \in X_\omega\}$$

as the *(relative) Lipschitz constant* of f. In complete analogy to non-stationary shift spaces, there is the following estimate for Birkhoff sums $S_n f_\omega := \sum_{k=0}^{n-1} f_{\theta^k\omega} \circ T^k_\omega$ (see also [**18**, p. 80]). For $n \leq m$, $x, y \in [a]_\omega$ with $a \in \mathscr{W}^m_\omega$, and a locally $(r, m-n+1)$-Hölder continuous function f, we have

$$|S_n f_\omega(x) - S_n f_\omega(y)| \leq r^{m-n} \sum_{k=1}^{\infty} \kappa(\theta^{n-k}\omega) r^k.$$

It follows from an application of the ergodic theorem that $\lim_{k\to\infty} k^{-1} \log \kappa(\theta^{n-k}\omega) = 0$ a.s. Hence, the radius of convergence of the power series on the right-hand side is equal to 1 and, in particular, it follows from $r < 1$ that the right-hand side is finite a.s. (see, e.g., [**6**]). However, if also $\int \kappa \, dP < \infty$, e.g. if condition (H) below holds, then monotone

https://doi.org/10.1017/etds.2015.61 Published online by Cambridge University Press

convergence and θ-invariance of P imply that

$$\int \sum_{k=1}^{\infty} \kappa(\theta^{n-k}\omega) r^k \, dP(\omega) = \sum_{k=1}^{\infty} r^k \int \kappa \, dP = \frac{r}{1-r} \int \kappa \, dP < \infty.$$

In particular, for a given locally Hölder continuous function φ with index less than or equal to $(m - n + 1)$, we hence have that

$$(B_{\theta^n\omega})^{-1} \le (B_{\theta^n\omega})^{-r^{m-n}} \le e^{S_n\varphi_\omega(x) - S_n\varphi_\omega(y)} \le (B_{\theta^n\omega})^{r^{m-n}} \le B_{\theta^n\omega},$$
$$\int \log B_\omega \, dP < \infty \quad \text{where } B_\omega := \exp \sum_{k=1}^{\infty} \kappa(\theta^{-k}\omega) r^k. \tag{10}$$

After these considerations, we are now in position to give the definition of the systems under consideration. That is, we consider $(X, T, \Omega, P, \theta)$ with the b.i.p. property and a potential function φ satisfying the following assumptions (H) and (S) on Hölder continuity and summability, respectively.

(H) The potential φ is locally Hölder continuous with index 2, and the associated random variable κ satisfies $\int \kappa \, dP < \infty$. Furthermore, $V_1^\omega(\varphi) < \infty$ for a.e. $\omega \in \theta^{-1}(\Omega_{\mathrm{bi}} \cup \Omega_{\mathrm{bp}})$.

(S) $\int |\log \mathscr{L}_{\theta^{-1}\omega}(1)| \, dP(\omega) < \infty$.

If $\mathscr{L}_\omega(1) = 1$ a.s., then the above condition (H) can be replaced by (Ĥ) below, which differs only by integrability of the Hölder constant.

(Ĥ) The potential φ is locally Hölder continuous with index 2, and $V_1^\omega(\varphi) < \infty$ for a.e. $\omega \in \theta^{-1}(\Omega_{\mathrm{bi}} \cup \Omega_{\mathrm{bp}})$.

Note that by definition of local Hölder continuity, condition (Ĥ) implies that the associated random variable κ satisfies $\int \log \kappa \, dP < \infty$. The reason behind that slightly more general integrability condition stems from the random variable λ in Theorem 4.1 (see also equation (2)). If $\mathscr{L}_\omega(1) = 1$, then automatically, $\lambda = 1$ a.s. However, if $\mathscr{L}_\omega(1) \neq 1$ on a set of positive measure, then the first step in the construction of λ is the construction of the pressure $P_G(\varphi)$ (as defined below), which depends on the almost subadditive ergodic theorem and therefore requires that $\log B_\omega$ is integrable (see [**6**, **18**]).

3.1. *Contraction of Vasershtein distances.* In order to apply Lemma 2.1, we now adapt the relevant variables and deduce their finiteness from the structure of (X, T), where we assume that (X, T) satisfies the b.i.p. property. For $\beta \in (0, 1)$, r given by the Hölder continuity of $\{\varphi_\omega : \omega \in \Omega\}$ and B_ω as in (10), define

$$\alpha_\omega := B_\omega/\beta, \quad \overline{d}_\omega(x, y) := \min\{1, \alpha_\omega d_r(x, y)\},$$

and let n_ω be an $\mathbb{N}$-valued random variable such that P-a.s.,

$$n_\omega \ge \lfloor -\log \alpha_\omega / \log r \rfloor + 1.$$

Furthermore, assume that $\mathbf{o}_\omega$ is an $\mathbb{N}$-valued random variable such that $\mathbf{o}_\omega \in \mathscr{W}_\omega^1$ (e.g. set $\mathbf{o}_\omega := \min(\mathscr{W}_\omega^1)$). Note that, since (X, T) is topologically mixing and θ is ergodic,

$$m_\omega := \min(\{n \ge 1 : \theta^n\omega \in \Omega_{\mathrm{bp}}, (A_\omega A_{\theta\omega} \cdots A_{\theta^{n-1}\omega})_{\mathbf{o}_\omega j} \ge 1 \; \forall j \in \mathscr{J} \cap \mathscr{W}^1_{\theta^{n-1}\omega}\})$$

is a.s. finite. Moreover, the big preimage property gives rise to the following argument. For each $x \in X_{\theta^{m_\omega}(\omega)}$, there exists $u(x) = (u_0 \ldots u_{m_\omega - 1}) \in \mathscr{W}^{m_\omega}$ such that $u_0 = \mathbf{o}_\omega$, $u_{m_\omega - 1} \in \mathscr{J}$ and $x \in T_\omega^{m_\omega}([u(x)]_\omega)$. By choosing $u(x)$ minimal according to the lexicographic order,

https://doi.org/10.1017/etds.2015.61 Published online by Cambridge University Press

we obtain a finite and measurable family $\{u(x) : x \in X_{\theta^{m_\omega}(\omega)}\}$. In particular,

$$C_\omega := \inf_{x \in X_{\theta^{m_\omega}(\omega)}} e^{S_{m_\omega}\varphi_\omega(\tau_{u(x)}x)} > 0$$

$$\text{and} \quad t_\omega := \max\left\{\beta,\, 1 - \frac{(1 - r^{n_\omega}\alpha_\omega)C_{\theta^{n_\omega}(\omega)}}{B_{\theta^{n_\omega}(\omega)}}\right\} \in (0,\, 1) \quad P\text{-a.s.} \tag{11}$$

For $\omega \in \Omega$, we refer to W_ω as the Vasershtein distance with respect to the metric $\overline{d}_\omega$ and to $\overline{D}_\omega(f)$ as the Lipschitz constant with respect to $\overline{d}_\omega$. In the context of random topological Markov chains, Lemma 2.1 is as follows.

LEMMA 3.1. *Assume that ($\hat{H}$) holds, that $f = \{f_\omega\}$ is Hölder continuous with respect to the same parameter as φ, and that $\mu = \{\mu_\omega\}$ and $\nu = \{\nu_\omega\}$ are random probability measures. For a.e. $\omega \in \Omega$, with $k_\omega := n_\omega + m_{\theta^{n_\omega}(\omega)}$, we then have that:*

(1) $\overline{D}_{\theta^{n_\omega}(\omega)}(\mathscr{L}_\omega^{n_\omega}(f)) \le \overline{D}_\omega(f)$;

(2) $\overline{D}_{\theta^{k_\omega}(\omega)}(\mathscr{L}_\omega^{k_\omega}(f)) \le t_\omega \cdot \overline{D}_\omega(f)$;

(3) $W_\omega((\mathscr{L}^{n_\omega})^*(\mu_{\theta^{n_\omega}(\omega)}), (\mathscr{L}^{n_\omega})^*(\nu_{\theta^{n_\omega}(\omega)})) \le W_{\theta^{n_\omega}(\omega)}(\mu_{\theta^{n_\omega}(\omega)}, \nu_{\theta^{n_\omega}(\omega)})$;

(4) $W_\omega((\mathscr{L}^{k_\omega})^*(\mu_{\theta^{k_\omega}(\omega)}), (\mathscr{L}^{k_\omega})^*(\nu_{\theta^{k_\omega}(\omega)})) \le t_\omega \cdot W_{\theta^{k_\omega}(\omega)}(\mu_{\theta^{k_\omega}(\omega)}, \nu_{\theta^{k_\omega}(\omega)})$.

Remark 1. The pathwise character of Lemma 2.1 admits in contrast to spectral theoretic methods like in [**7**, **12**, **20**] an immediate application in a fibred setting. Moreover, it is worth noting that a recent result of Zhang in [**22**] establishes a fibrewise Monge–Kantorovich duality, which would allow, even though not necessary for the proof, choosing the couplings $Q_{x,y}$ and Q in step (3) in a measurable way.

As an application of Lemma 3.1, that is, through construction of n_ω and m_ω, we obtain the key observation of this note, giving rise to exponential decay of the Vasershtein distance W with respect to the initial shift metric d_r. For completeness, the dual statement on decay of Lipschitz constants also is included.

COROLLARY 3.1. *Assume that (X, T, φ) satisfies the b.i.p. property, that ($\hat{H}$) holds and that $\mathscr{L}_\omega(1) = 1$ a.s. Then there exist constants $t \in (0, 1)$, $c \in (0, \infty)$ and random sequences $(k_n(\omega) : n \in \mathbb{N})$ and $(l_n(\omega) : n \in \mathbb{N})$ such that:*

(1) *for random probability measures μ, ν and $n \in \mathbb{N}$, with $\mu_\omega^n := (\mathscr{L}_\omega^n)^*(\mu_{\theta^n\omega})$,*

$$W(\mu^{k_n(\omega)}_{\theta^{-k_n(\omega)}(\omega)}, \nu^{k_n(\omega)}_{\theta^{-k_n(\omega)}(\omega)}) \le 2B_\omega \cdot t^n W(\mu_\omega, \nu_\omega),$$

$$W(\mu_\omega^{l_n(\omega)}, \nu_\omega^{l_n(\omega)}) \le c \cdot t^n W(\mu_{\theta^{l_n(\omega)}(\omega)}, \nu_{\theta^{l_n(\omega)}(\omega)});$$

(2) *for a fibrewise Lipschitz continuous function f and $n \in \mathbb{N}$,*

$$D(\mathscr{L}^{k_n(\omega)}_{\theta^{-k_n(\omega)}(\omega)}(f)) \le 2B_\omega \cdot t^n D(f_{\theta^{-k_n(\omega)}(\omega)}),$$

$$D(\mathscr{L}_\omega^{l_n(\omega)}(f)) \le c \cdot t^n D(f_\omega).$$

If B_ω is uniformly bounded, then the above are satisfied for $c := 2\,\text{ess-sup}_{\omega\in\Omega}(B_\omega)$.

Proof. We begin with the construction of $(l_n(\omega))$. In order to apply Lemma 3.1, choose $\beta = 1/2$ and $C > 0$ and $B \ge 1$ such that $P(\Omega_{B,C}) > 0$, where $\Omega_{B,C} := \{\omega \in \Omega : B_\omega \le B,\, C_\omega \ge C\}$. Note that this implies that $\alpha_\omega = 2B_\omega$. For

$$n_\omega := \min(\{n \in \mathbb{N} : \theta^n(\omega) \in \Omega_{B,C},\, n \ge \lfloor -\log(2\alpha_\omega)/\log r \rfloor + 1\}),$$

https://doi.org/10.1017/etds.2015.61 Published online by Cambridge University Press

it follows that $t := 1 - C/2B \geq t_\omega$ for P-a.e. ω. The sequence $l_n(\omega)$ is now defined as follows. For $n = 0$, set $l_0(\omega) := n_\omega$ and define l_n by

$$l_n(\omega) := l_{n-1}(\omega) + m_{\theta^{l_{n-1}(\omega)}(\omega)} + n_{\theta^{l^*(n-1,\omega)}(\omega)},$$

where $l^*(n-1, \omega) = l_{n-1}(\omega) + m_{\theta^{l_{n-1}(\omega)}(\omega)}$. Hence, $\theta^{l_n(\omega)}(\omega) \in \Omega_{B,C}$ and, in particular, $\alpha_{\theta^{l_n(\omega)}(\omega)} \leq 2B := c$. In order to obtain $(l_n(\omega))$, we adapt the above to negative powers of θ. That is, with $\widetilde{\Omega}_{B,C} := \{\theta^{m_\omega} : \omega \in \Omega_{B,C}\}$, define

$$\begin{aligned}
\widetilde{n}_\omega &:= \min(\{n \in \mathbb{N} : \theta^{-n}(\omega) \in \widetilde{\Omega}_{B,C},\, n \geq \lfloor -\log(2\alpha_{\theta^{-n}(\omega)})/\log r \rfloor + 1\}),\\
\widetilde{m}_\omega &:= \min(\{n : \theta^{-n}\omega \in \Omega_{B,C},\, n = m_{\theta^{-n}(\omega)}\}) \quad \text{for } \omega \in \widetilde{\Omega}_{B,C},\\
k_0(\omega) &:= \widetilde{n}_\omega + \widetilde{m}_{\theta^{-\widetilde{n}_\omega}(\omega)}, \quad k_n(\omega) := k_{n-1}(\omega) + \widetilde{n}_{\theta^{-k_{n-1}(\omega)}(\omega)} + \widetilde{m}_{\theta^{-k^*(n-1,\omega)}(\omega)},\\
&\quad \text{with} \quad k^*(n-1, \omega) = k_{n-1}(\omega) + \widetilde{n}_{\theta^{-k_{n-1}(\omega)}(\omega)}.
\end{aligned}$$

The assertions of the corollary follow from Lemma 3.1 and $\alpha_\omega d \geq \overline{d}_\omega \geq d$. □

For the statement of the corollary, the exponential decay was formulated with respect to the sequences (k_n) and (l_n) since this approach allows us to explicitly construct the involved constants, which might be useful for questions concerning stochastic stability (as, e.g., in [**17**]). However, a slight modification of the construction allows us to obtain a more classical formulation of exponential decay. For ease of exposition, the result contains only the statement with respect to the Vasershtein distance.

COROLLARY 3.2. *There exists $s \in (0, 1)$ and a positive random variable c^*_ω such that, for each pair of random probability measures μ, ν and $n \in \mathbb{N}$, we have*

$$W(\mu^n_{\theta^{-n}(\omega)}, \nu^n_{\theta^{-n}(\omega)}) \leq c^*_\omega \cdot s^n W(\mu_\omega, \nu_\omega) \quad \textit{and} \quad W(\mu^n_\omega, \nu^n_\omega) \leq c^*_\omega \cdot s^n W(\mu_{\theta^n(\omega)}, \nu_{\theta^n(\omega)}).$$

Proof. Similar to the proof of Corollary 3.1, choose B, C, M such that

$$\Omega_{B,C,M} := \{\omega \in \Omega : B_\omega \leq B,\, C_\omega \geq C,\, m_\omega \leq M,\, B_{\theta^{m_\omega}\omega} \leq B\}$$

has positive measure. Furthermore, set $K := \lfloor -\log(2B)/\log r \rfloor + 1$ and let η^ω_k refer to the kth entrance time to $\Omega_{B,C,M}$, that is,

$$\eta^\omega_k := \min(\{n_k \in \mathbb{N} : \exists\, 0 < n_1 < \cdots < n_k \text{ s.t. } \theta^{n_i}(\omega) \in \Omega_{B,C,M} \text{ for } i = 1, \ldots, k\}).$$

Since $\eta^\omega_k \geq k$ for all $k \in \mathbb{N}$, it follows that for a.e. $\omega \in \Omega$, $n_{\theta^{m_\omega}(\omega)} := \eta^\omega_{M+K} - m_\omega \geq K$. Hence, Lemma 3.1 is applicable to each transition from $l_k(\omega)$ to $l_{k+1}(\omega)$, where $l_k(\omega)$ is defined by $l_k(\omega) := \eta^\omega_{(M+K)k} + m_{\theta^{\eta^\omega_{(M+K)k}}\omega}$. For $s = 1 - C/2B$ as in the proof above, it follows from the comparability of d and $\overline{d}_\omega$ that for a.e. ω,

$$W((\mathscr{L}^{l_k(\omega)-m_\omega}_{\theta^{m_\omega}(\omega)})^*\mu_{\theta^{l_k(\omega)}\omega}, (\mathscr{L}^{l_k(\omega)-m_\omega}_{\theta^{m_\omega}(\omega)})^*\nu_{\theta^{l_k(\omega)}\omega}) \leq 2Bs^k W(\mu_{\theta^{l_k(\omega)}\omega}, \nu_{\theta^{l_k(\omega)}\omega}).$$

https://doi.org/10.1017/etds.2015.61 Published online by Cambridge University Press

In order to obtain a rate with respect to n as in the statement of the corollary, choose k such that $l_k \le n < l_{k+1}$. By applying the ergodic theorem twice, it follows that

$$\frac{1}{n}\log s^k = \frac{l_k}{n}\frac{k}{l_k}\log s \xrightarrow{k,n\to\infty} \frac{1}{(M+K)P(\Omega_{B,C,M})}\log s.$$

Moreover, note that substituting k with $k+1$ does not change the limit. Hence, for each $t \in (0,1)$ with $\log t > (\log s)/((M+K)P(\Omega_{B,C,M}))$, there is a random variable c^* such that

$$W((\mathscr{L}_\omega^{\theta^n\omega})^*\mu_{\theta^n\omega}, (\mathscr{L}_\omega^{\theta^n\omega})^*\nu_{\theta^n\omega}) \le c_\omega^* t^n W(\mu_{\theta^n\omega}, \nu_{\theta^n\omega}).$$

The remaining assertion follows from the same arguments by considering the $(M+K)$th entrances to $\Omega_{B,C,M}$ with respect to θ^{-1} instead of θ. □

4. *The random version of Ruelle's theorem*

We now apply the above results in order to obtain exponential decay in Ruelle's theorem. We first recall the basic results from [**18**] for $(X, T, \Omega, P, \theta)$ with the b.i.p. property and conditions (H) and (S). For a given $a \in \mathscr{W}^1$ and a measurable family $\{\xi_\omega \in [a]_\omega : \omega \in \Omega\}$, the *nth local preimage function* is defined by

$$\mathscr{Z}_n^\omega(a) := \mathscr{L}_\omega^n(1_{[a]})(\xi_{\theta^n\omega}) \quad \text{for all } \omega \in \Omega_a \cap \theta^{-n}(\Omega_a).$$

The *relative Gurevič pressure* $P_G(\varphi)$ is then defined as follows. Due to the restriction on ω and n in the definition of the local preimage function, the classical definition of pressure has to be modified as a limit along a subsequence given by returns with respect to the base. More precisely, for $\Omega' \subset \Omega$ of positive measure and $\omega \in \Omega$, define $J_\omega(\Omega') := \{n \in \mathbb{N} : \theta^n\omega \in \Omega'\}$. For $N \in \mathbb{N}$ such that $\Omega^* := \{\omega \in \Omega_a : N_{aa}(\omega) \le N\}$ is a set of positive measure, the Gurevič pressure of (X, T, φ) is then defined by

$$P_G(\varphi) := \lim_{n\to\infty, n\in J_\omega(\Omega^*)} \frac{1}{n}\log \mathscr{Z}_n^\omega(a). \tag{12}$$

By [**6**, Theorem 3.2], the above limit exists, is finite, P-a.s. constant for $\omega \in \Omega_a$ and independent of the choices of a and N. Combining [**18**, Theorem 4.1] with Corollaries 3.1 and 3.2 above then allows us to obtain the main result of this note. In order to state the theorems, we define the following random Lipschitz norm with respect to d_r, where r is given by the Hölder parameter of φ. For $\omega \in \Omega$ and $f_\omega : X_\omega \to \mathbb{R}$ with $D_\omega(f) < \infty$, set

$$\|f_\omega\|_L^\omega := D_\omega(f_\omega) + \|f_\omega\|_\infty.$$

Furthermore, for the statement of the theorems, let c, c_ω^* be as in Corollaries 3.1 and 3.2 and, for a given random variable $\lambda : \Omega \to (0, \infty)$, define $\Lambda_k(\omega) := \prod_{i=0}^{k-1} \lambda_{\theta^i(\omega)}$.

THEOREM 4.1. *Assume that (X, T, φ) satisfies the b.i.p. property and* (H) *and* (S) *hold. Then there exist a random variable $\lambda : \Omega \to \mathbb{R}$ with $\int \log \lambda_\omega \, dP = P_G(\varphi)$, a measurable family of functions $\{h_\omega : \omega \in \Omega\}$ and a random probability measure $\{\mu_\omega : \omega \in \Omega\}$ such that the following hold.*

(1) *For a.e. $\omega \in \Omega$, $h_\omega : X_\omega \to \mathbb{R}$ is a strictly positive function satisfying $\mathscr{L}_\omega h_\omega = \lambda_\omega h_{\theta\omega}$ and $\int h_\omega \, d\mu_\omega = 1$.*

(2) *$\{\log h_\omega\}$ is Hölder continuous with Hölder constant of index 1 bounded by $B_\omega - 1$ and the same Hölder parameter as φ.*

https://doi.org/10.1017/etds.2015.61 Published online by Cambridge University Press

(3) *For a.e.* $\omega \in \Omega$, $\mathscr{L}_\omega^*(\mu_{\theta\omega}) = \lambda_\omega \mu_\omega$.
(4) *The probability measure given by* $h_\omega \, d\mu_\omega \, dP$ *is* T*-invariant and ergodic.*
(5) $\{\mu_\omega\}$ *is the unique random probability measure with* $\mathscr{L}_\omega^*(\mu_{\theta\omega}) = \lambda_\omega \mu_\omega$ *a.s. Furthermore,* $\{h_\omega\}$ *is, up to scalar multiplication, the unique positive, measurable and non-trivial function with* $\mathscr{L}_\omega h_\omega = \lambda_\omega h_{\theta\omega}$ *a.s.*
(6) *There exists* $t \in (0, 1)$*, a positive random variable* K *and a random sequence* $(l_n(\omega) : n \in \mathbb{N})$ *such that for each fibrewise Lipschitz continuous function* $f = \{f_\omega\}$, $n \in \mathbb{N}$ *and a.e.* $\omega \in \Omega$,

$$\left\| \frac{\mathscr{L}_\omega^{l_n(\omega)}(f_\omega)}{\Lambda_{l_n}(\omega) h_{\theta^{l_n}(\omega)}} - \int f \, d\mu_\omega \right\|_L^{\theta^{l_n}\omega} \le c K_\omega \cdot t^n \|f\|_L^\omega.$$

(7) *There exists* $s \in (0, 1)$ *such that for each fibrewise Lipschitz continuous function* $f = \{f_\omega\}$, $n \in \mathbb{N}$ *and a.e.* $\omega \in \Omega$,

$$\left\| \frac{\mathscr{L}_\omega^n(f)}{\Lambda_n(\omega) h_{\theta^n(\omega)}} - \int f \, d\mu_\omega \right\|_L^{\theta^n\omega} \le c_\omega^* K_\omega \cdot s^n \|f\|_L^\omega.$$

The proof of the theorem makes use of a normalized version of Ruelle's operator, which is necessary in order to obtain contraction of the Vasershtein metric. That is, with $\{h_\omega\}$ and λ as in the theorem, let $\tilde{\mathscr{L}}_\omega$ refer to the Ruelle operator with respect to the potential

$$\tilde{\varphi}_\omega(x) := \varphi_\omega(x) + \log h_\omega(x) - \log h_{\theta\omega}(T_\omega(x)) - \log \lambda_\omega.$$

As an application of Corollaries 3.1 and 3.2 we obtain the following theorem.

THEOREM 4.2. *Assume that* (X, T, ϕ) *satisfies the b.i.p. property and* (H) *and* (S) *hold. Then there exist a random probability measure* $\{\nu_\omega : \omega \in \Omega\}$*, constants* $t \in (0, 1)$*,* $c \in (0, \infty)$ *and random sequences* $(k_n(\omega) : n \in \mathbb{N})$ *and* $(l_n(\omega) : n \in \mathbb{N})$ *such that for each fibrewise Lipschitz continuous function* $f = \{f_\omega\}$, $n \in \mathbb{N}$ *and a.e.* $\omega \in \Omega$,

$$\left\| \tilde{\mathscr{L}}_\omega^{l_n(\omega)}(f) - \int f \, d\nu_\omega \right\|_L^{\theta^{l_n}\omega} \le 2c \cdot t^n D_\omega(f) \quad \text{and}$$

$$\left\| \tilde{\mathscr{L}}_{\theta^{-k_n(\omega)}(\omega)}^{k_n(\omega)}(f) - \int f \, d\nu_{\theta^{-k_n(\omega)}(\omega)} \right\|_L^\omega \le 4B_\omega \cdot t^n D_{\theta^{-k_n(\omega)}(\omega)}(f).$$

Furthermore, $\{\nu_\omega : \omega \in \Omega\}$ *is the unique measure with* $\tilde{\mathscr{L}}_\omega^*(\nu_{\theta\omega}) = \nu_\omega$.

Remark 2. Note that the second estimate in Theorem 4.2 also could be reformulated in terms of $\mathscr{L}_\omega$ through identity (15). Namely, it immediately follows from the proof that a.s., for each Lipschitz continuous function $\{f_\omega\}$,

$$\left\| \frac{\mathscr{L}_{\theta^{-n}(\omega)}^n(f)}{\Lambda_n(\theta^{-n}(\omega)) h_{\theta^{-n}(\omega)}} - \int f \, d\mu_{\theta^{-n}(\omega)} \right\|_L^\omega \le c_\omega^* K_{\theta^{-n}(\omega)} \cdot s^n \|f\|_L^{\theta^{-n}(\omega)},$$

where s and K are given by Theorem 4.1. Moreover, there exists a random sequence $(k_n(\omega) : n \in \mathbb{N})$ such that, with t as in Theorem 4.1 and using $\alpha(\omega) := 2B_\omega$,

$$\left\| \frac{\mathscr{L}_{\theta^{-k_n}(\omega)}^{k_n}(f)}{\Lambda_{k_n}(\theta^{-k_n}(\omega)) h_{\theta^{-k_n}(\omega)}} - \int f \, d\mu_{\theta^{-k_n}(\omega)} \right\|_L^\omega \le 2B_\omega K_{\theta^{-k_n}(\omega)} \cdot t^n \|f\|_L^{\theta^{-k_n}(\omega)}.$$

https://doi.org/10.1017/etds.2015.61 Published online by Cambridge University Press

Remark 3. The advantage in stating the exponential decay of correlations with respect to subsequences is that the constants are explicitly given, which could give rise to a refined analysis of stochastic stability of randomly perturbed intermittent maps as studied, e.g. in [**17**]. Since the contraction in (6) and (7) of Theorem 4.1 stems from returns to a set with bounded parameters, c can be chosen to be any value bigger than ess-inf$(2B_\omega)$, even though the choice affects the sequences (k_n) and (l_n), which are essentially defined through consecutive visits to Ω_{bp} and a subset of Ω with bounded parameters. Furthermore, observe that the random variable K_ω does not depend on the construction of (k_n) and (l_n). Namely, K_ω depends only on B_ω and $\{h_\omega\}$, that is, $K_\omega = 2\|1/h_\omega\|_\infty \cdot \max\{\|h_\omega\|_\infty \|1/h_\omega\|_\infty - 1,\, B_\omega - 1,\, 1\}$ as shown in (18) below.

Remark 4. We remark that there is an error in the proof of [**18**, Theorem 4.1], which states that $\mathscr{L}^*_\omega(\mu_{\theta\omega}) = \lambda_\omega \mu_\omega$ and that $P_\omega(s)/P_{\theta\omega}(s) \to \lambda_\omega$ converges a.s. The proof of the first statement is correct and follows the lines of the proof of [**6**, Proposition 6.3]. However, since convergence in the narrow topology does not imply weak convergence almost surely, the proof of the second statement is not correct (see also the corrigendum in [**19**]). However, combining $\lambda_\omega = \int \mathscr{L}_\omega(1)\, d\mu_{\theta\omega}$ with (7) in Theorem 4.1 implies that there exists a positive random variable L such that, for a.e. ω,

$$\|\mathscr{L}^{n+1}_\omega \varphi(1)/\mathscr{L}^n_{\theta\omega}\varphi(1) - \lambda_\omega\|_\infty \le L_\omega s^n,$$

which is a significantly stronger statement than the convergence of $P_\omega(s)/P_{\theta\omega}(s)$ to λ_ω.

4.1. *Application to random positive matrices.* For illustration, we give an application of Theorem 4.1 to products of random positive matrices. Observe that a positive matrix always can be considered as a Ruelle operator with respect to a potential that is constant on cylinders of length 2. Or in other words, the potential is locally Hölder continuous with index 2 and $\kappa = 0$, $B = 1$. As in [**18**], we consider random matrices $A = \{A_\omega : \omega \in \Omega\}$ with $A_\omega = (p^\omega_{ij},\, i \in \mathscr{W}^1_\omega,\, j \in \mathscr{W}^1_{\theta\omega})$ and $p_{ij} \ge 0$ a.s. We then refer to A as a summable random matrix with the b.i.p. property if:

(1) the signum of A defines a random topological Markov chain with the b.i.p. property;
(2) for a.e. $\omega \in \theta^{-1}(\Omega_{\mathrm{bi}} \cup \Omega_{\mathrm{bp}})$, we have

$$\sup\{p^\omega_{ij}/p^\omega_{ik} : i \in \mathscr{W}^1_\omega,\, j, k \in \mathscr{W}^1_{\theta\omega},\, p^\omega_{ik} \ne 0\} < \infty;$$

(3) the random variable $\omega \mapsto \sup_{j \in \mathscr{W}^1_{\theta\omega}} |\log \sum_{i \in \mathscr{W}^1_\omega} p^\omega_{ij}|$ is in $L^1(P)$.

As in [**18**], it follows that Theorem 4.1 is applicable and that the resulting eigenfunction and conformal measure are in fact vectors. Hence, for each summable random matrix A with the b.i.p. property, there exist a positive random variable λ and strictly positive random vectors $h = \{h^\omega \in \mathbb{R}^{\ell_\omega - 1} : \omega \in \Omega\}$ and $\mu = \{\mu^\omega \in \mathbb{R}^{\ell_{\theta\omega} - 1} : \omega \in \Omega\}$ such that, for a.e. $\omega \in \Omega$, we have $(h^\omega)^t A_\omega = \lambda_\omega (h^{\theta\omega})^t$, $A_\omega \mu^{\theta\omega} = \lambda_\omega \mu^\omega$. By applying (7) of Theorem 4.1 to the ith unit vector $e_i := (\delta_{ij} : j \in \mathscr{W}^1_\omega)$, we obtain

$$\sup_{i \in \mathscr{W}^1_\omega,\, j \in \mathscr{W}^1_{\theta^n\omega}} \left| \frac{(A^n_\omega)_{ij}}{\Lambda_n(\omega) h^{\theta^n\omega}_j} - \mu^\omega_i \right| \le c^*_\omega K_\omega \cdot s^n,$$

https://doi.org/10.1017/etds.2015.61 Published online by Cambridge University Press

with δ_{ij} referring to Kronecker's δ function and $(\cdot)_{ij}$ to the coordinate (i, j) of a matrix. As can be seen easily, the above estimate implies exponential convergence of $\Lambda_n(\omega)^{-1}A_\omega^n$ in the sup-norm for matrices, that is,

$$\|\Lambda_n(\omega)^{-1}A_\omega^n - \mu^\omega \cdot (h^{\theta^n\omega})^t\|_\infty \le c_\omega^* K_\omega \left(\sup_j h_j^{\theta^n\omega}\right) s^n.$$

Let $\tilde{A}$ be the random matrix with constants $(\tilde{A}_\omega)_{i,j} := (A_\omega)_{i,j}h_i^\omega/(\lambda_\omega h_j^{\theta\omega})$. Since the matrix corresponds to Ruelle's operator with a potential that is constant on cylinders of length 2, we have that $B_\omega = 1$ for all $\omega \in \Omega$. By Theorem 4.2, it follows that there exists a positive random vector $\{\nu^\omega\}$ with $\sum_{i\in\mathscr{W}_\omega^1} \nu_i^\omega = 1$ a.s. such that, for all $n \in \mathbb{N}$,

$$\sup_{\substack{i\in\mathscr{W}_\omega^1,\\ j\in\mathscr{W}_{\theta^{l_n}(\omega)}^1}} |(\tilde{A}_\omega^{l_n(\omega)})_{ij} - \nu_i^\omega| \le 4\cdot t^n, \qquad \sup_{\substack{i\in\mathscr{W}_{\theta^{-k_n}(\omega)}^1,\\ j\in\mathscr{W}_\omega^1}} |(\tilde{A}_{\theta^{-k_n}(\omega)}^{k_n(\omega)})_{ij} - \nu_i^{\theta^{-k_n}(\omega)}| \le 4\cdot t^n.$$

To obtain control on the sequences l_n and k_n and the parameter t, we have to impose further conditions on $\tilde{A}_\omega$. For example, if $(\tilde{A}_\omega)_{\mathbf{o}j} \ge C > 0$ for all $\omega \in \theta^{-1}(\Omega_{\mathrm{bp}})$ and $j \in \mathscr{W}_{\theta\omega}^1$, then $t = 1 - C/2$. Moreover, by setting $r < 1/2$, it follows that $\lfloor -\log\alpha_\omega/\log r\rfloor = 0$. By the proof of Corollary 3.1, the above is satisfied for the following sequences (l_n) and (k_n): for $\omega \in \Omega$, set $l_1(\omega) := \min\{n \ge 2 : \theta^n(\omega) \in \Omega_{\mathrm{bp}}\}$ and by induction, for $n \in \mathbb{N}$,

$$l_n(\omega) := l_{n-1}(\omega) + l_1(\theta^{l_{n-1}(\omega)}(\omega)).$$

Due to contraction only along passages from $\theta^{-1}(\Omega_{\mathrm{bp}})$ to Ω_{bp}, there is a slight asymmetry in the construction of (k_n). Set $k_0(\omega) := \min\{n \ge 1 : \theta^{-n}(\omega) \in \Omega_{\mathrm{bp}}\}$ and $u(\omega) := \min\{n \ge 2 : \theta^{-n}(\omega) \in \Omega_{\mathrm{bp}}\}$. Then (k_n) is defined by, for $n \in \mathbb{N}$,

$$k_n(\omega) := u(\theta^{-k_{n-1}(\omega)}(\omega)) + k_{n-1}(\omega).$$

For example, if $\Omega_{\mathrm{bp}} = \Omega$, then (l_n) and (k_n) are the sequence $(2n : n \in \mathbb{N})$. In this case, an argument as in Subsection 4.3 below allows us to substitute for $(2n : n \in \mathbb{N})$ with the sequence $(n : n \in \mathbb{N})$.

These results are related to [**15**, Problem 5.7]. Suppose that A is a (random) positive matrix such that conditions (1) and (2) above are satisfied and that the transpose is almost surely a stochastic matrix, that is, $1 \cdot A_\omega = 1$. Observe that this implies that condition (3) is satisfied, that $\lambda_\omega = 1$ a.s. and that h_ω is almost surely equal to the function 1. Hence, $A = \tilde{A}$ and, by the above, the iterates of A are contractions of the Vasershtein metric along subsequences, which then implies exponential decay. However, this contraction property is a milder condition than $\delta^* = 0$ in [**15**], which essentially means eventual contraction of the bounded variation norm.

4.2. *Application to decay of correlations.* For (X, T) with the b.i.p. property, $(\hat{\mathrm{H}})$ and $\mathscr{L}_\omega(1) = 1$ a.s., Theorem 4.2 also allows us to deduce the pathwise exponential decay of correlations using the fundamental identity

$$\int f_\omega \cdot g_{\theta^n\omega} \circ T_\omega^n \, d\mu_\omega = \int \mathscr{L}_\omega^n(f_\omega) \cdot g_{\theta^n\omega} \, d\mu_{\theta^n\omega},$$

for a.e. $\omega \in \Omega$, f Lipschitz and g with $\int |g_{\theta^n\omega}| \, d\mu_{\theta^n\omega} < \infty$. Now assume that $\int f_\omega \, d\mu_\omega = 0$. Then Theorem 4.2 implies that $\|\mathscr{L}_\omega^{l_n(\omega)}(f_\omega) \, d\mu_\omega\|_\infty$ and

https://doi.org/10.1017/etds.2015.61 Published online by Cambridge University Press

$\|\mathscr{L}^{k_n(\omega)}_{\theta^{-k_n(\omega)}\omega}(f_{\theta^{-k_n(\omega)}\omega})\,d\mu_{\theta^{-k_n(\omega)}\omega}\|_\infty$ converge to 0 exponentially fast. Hence, for $f=\{f_\omega\}$ Lipschitz continuous with $\int f\,d\mu_\omega=0$ a.s. and $g=\{g_\omega\}$ with $\int|g|\,d\mu_\omega<\infty$ a.s., we obtain the following estimates for the decay of correlations:

$$\left|\int f\cdot g\circ T^{l_n(\omega)}_\omega\,d\mu_\omega\right|\le 2c\cdot t^n D_\omega(f)\int|g|\,d\mu_{\theta^{l_n(\omega)}\omega},$$
$$\left|\int f\cdot g\circ T^{k_n(\omega)}_{\theta^{-k_n(\omega)}\omega}\,d\mu_{\theta^{-k_n(\omega)}\omega}\right|\le 4B_\omega\cdot t^n D_{\theta^{-k_n(\omega)}\omega}(f)\int|g|\,d\mu_\omega.$$

4.3. *Application to ψ-mixing coefficients.* As a corollary of Theorem 4.2, we obtain a mixing property by specifying the ψ-mixing coefficients known from probability theory. In case of a random topological Markov chain, these are defined by

$$\psi_n(\omega):=\sup\frac{\mu_{\theta^{-k}\omega}([a]\cap T^{-k-n}_{\theta^{-k}\omega}(A))-\mu_{\theta^{-k}\omega}([a])\mu_{\theta^n\omega}(A)}{\mu_{\theta^{-k}\omega}([a])\mu_{\theta^n\omega}(A)},$$

where the supremum is taken with respect to all $k\in\mathbb{N}$, $a\in\mathscr{W}^k_{\theta^{-k}\omega}$ and $A\subset X_{\theta^n\omega}$ measurable such that $\mu_{\theta^{-k}\omega}([a])$, $\mu_{\theta^n\omega}(A)>0$. Observe that, for $a\in\mathscr{W}^k$ with $P(\Omega_a)>0$ and

$$f^\omega_a(x):=\mathbf{1}_{\theta^k(\Omega_a)}(\omega)e^{S_k\varphi_{\theta^{-k}\omega}\circ\tau_a(x)},$$

the estimate (10) implies that f^ω_a is Lipschitz continuous and, in particular, that

$$D_\omega(f^\omega_a)\le\frac{B^2_\omega}{\mu_\omega(T_{\theta^{-k}\omega}([a]))}\mu_{\theta^{-k}\omega}([a]).$$

Now assume that (X,T) satisfies the b.i.p. property, ($\hat{\mathrm{H}}$) and $\mathscr{L}_\omega(1)=1$ a.s. Then $\{\mu_\omega\}$ is a random invariant measure and Theorem 4.2 implies that, for $n\in\mathbb{N}$ and $A\subset X_{\theta^{l_n(\omega)}\omega}$ measurable,

$$\begin{aligned}
&|\mu_{\theta^{-k}\omega}([a]\cap T^{-k-l_n(\omega)}_{\theta^{-k}\omega}(A))-\mu_{\theta^{-k}\omega}([a])\mu_{\theta^{l_n(\omega)}\omega}(A)|\\
&\quad=\left|\int f^\omega_a\mathbf{1}_A\circ T^{l_n(\omega)}_\omega\,d\mu_\omega-\int\mu_{\theta^{-k}\omega}([a])\mathbf{1}_A\circ T^{l_n(\omega)}_\omega\,d\mu_\omega\right|\\
&\quad=\left|\int(\mathscr{L}^{l_n(\omega)}_\omega(f^\omega_a)-\mu_{\theta^{-k}\omega}([a]))\mathbf{1}_A\,d\mu_{\theta^{l_n(\omega)}\omega}\right|\le 2c\mu_{\theta^{l_n(\omega)}\omega}(A)D_\omega(f^\omega_a)\\
&\quad\le\frac{2cB^2_\omega}{\mu_\omega(T_{\theta^{-k}\omega}([a]))}t^n\mu_{\theta^{-k}\omega}([a])\mu_{\theta^{l_n(\omega)}\omega}(A).
\end{aligned}$$

In particular, if $\omega\in\Omega_{\mathrm{bi}}$, then $\mu_\omega(T_{\theta^{-k}\omega}([a]))$ is bounded from below. Hence, there exists $\Omega'\subset\Omega_{\mathrm{bi}}$ of positive measure and $C>0$ such that, for all $\omega\in\Omega'$,

$$\psi_{l_n}(\omega)\le Ct^n. \tag{13}$$

Furthermore, recall that the random sequence l_n is constructed through not necessarily first returns as follows. Let $\Omega_{B,C}$ be as in the proof Corollary 3.1, that is, B_ω and C_ω (see (11)) are uniformly bounded on $\Omega_{B,C}$. The sequence (l_n) then has to be chosen such that $\theta^{l_n(\omega)}(\omega)\in\Omega_{B,C}$ for all $n\in\mathbb{N}$ and such that the difference between two consecutive elements of (l_n) is bounded from below by the random variable $L(\omega):=m_\omega+n_{\theta^{m_\omega}(\omega)}$

https://doi.org/10.1017/etds.2015.61 Published online by Cambridge University Press

(see Lemma 3.1), that is,

$$l_n(\omega) - l_{n-1}(\omega) \geq L(\theta^{l_{n-1}(\omega)}(\omega)).$$

In the following, (l_n) is specified in case of a random full shift, B_ω is uniformly bounded from above and $\varphi_\omega|_{[\mathbf{o}_\omega]}$ is uniformly bounded from below. Under these conditions, $K := \text{ess-sup}\{m_\omega + n_{\theta^{m_\omega}(\omega)}\} < \infty$ and, in particular, (13) holds for $(l_n) = (Kn + k : n \in \mathbb{N})$, for each $k = 0, \ldots, K-1$. Hence, we have that, for all $\omega \in \Omega'$ and with $\tilde{C} := C/t$ and $\tilde{t} := \sqrt[K]{t}$ and for all $n \geq K$,

$$\psi_n(\omega) \leq \tilde{C}\tilde{t}^n.$$

4.4. *Application to equilibrium states.* Under the assumptions of Theorem 4.1, it was shown in [**6**] that a variational principle holds. That is,

$$P_G(\varphi) = \sup\left\{h_m^{(\mathrm{r})}(T) + \int \varphi \, dm_\omega \, dP : m \in \mathscr{M}_\theta(T)\right\},$$

where $h_m^{(\mathrm{r})}(T)$ refers to the fibre entropy as defined in [**14**] and $\mathscr{M}_\theta(T)$ refers to the random invariant probability measures, that is, to those random probability measures $m = \{m_\omega\}$ such that $m_\omega \circ T_\omega^{-1} = m_{\theta\omega}$ a.s. For the random probability measure $\nu = \{\nu_\omega\}$ defined by $d\nu_\omega = h_\omega \, d\mu_\omega$, with $\{h_\omega\}$ and $\{\mu_\omega\}$ given by Theorem 4.1, it then follows that

$$P_G(\varphi) = h_\nu^{(\mathrm{r})}(T) + \int \varphi_\omega \, d\nu_\omega \, dP,$$

that is, $\{\nu_\omega\}$ is an equilibrium state. In order to prove this assertion, recall from [**14**] that

$$h_m^{(\mathrm{r})}(T) = -\lim_{n\to\infty} \sum_{a \in \mathscr{W}_\omega^n} m_\omega([a]) \log m_\omega([a]).$$

By the Gibbs property in [**18**, Remark 4.2] (as in (14) below), there exists a subset Ω' of positive measure of Ω_{bi} and $F > 0$ such that $\mu_\omega([a]_\omega) = F^{\pm 1} \exp(S_n\varphi_\omega(x))/\Lambda_n(\omega)$ for all $n \in \mathbb{N}$ with $\theta^n\omega \in \Omega'$, $a \in \mathscr{W}_\omega^n$ and $x \in [a]_\omega$. For $d\nu_\omega = h_\omega \, d\mu_\omega$, we hence have by invariance of $\{\nu_\omega\}$, Hölder continuity of h_ω and Birkhoff's theorem that

$$\begin{aligned}
h_\mu^{(\mathrm{r})}(T) &= \lim_{n\to\infty, \theta^n\omega\in\Omega'} -\frac{1}{n} \sum_{a\in\mathscr{W}_\omega^n} \nu_\omega([a]_\omega) \log \nu_\omega([a]_\omega) \\
&= \lim_{n\to\infty, \theta^n\omega\in\Omega'} -\frac{1}{n} \int_{X_\omega} \log h_\omega + S_n\varphi_\omega - \log \Lambda_n(\omega) \, d\nu_\omega \\
&= \lim_{n\to\infty, \theta^n\omega\in\Omega'} \frac{1}{n}\left(\int_{X_\omega} \log h_\omega \, d\nu_\omega + \sum_{k=0}^{n-1}\left(\log \lambda_{\theta^k\omega} - \int_{X_\omega} \varphi_{\theta^k\omega} \circ T_\omega^k \, d\nu_\omega\right)\right) \\
&= P_G(\varphi) - \int \varphi_\omega \, d\nu_\omega \, dP.
\end{aligned}$$

Hence, $\{\nu_\omega\}$ is an equilibrium state for (X, T).

5. *Proofs of Theorems 4.1 and 4.2*

This section is exclusively devoted to the proofs of Theorems 4.1 and 4.2. Observe that parts (3) and (4) of Theorem 4.1 are as in [**18**, Theorem 4.2] and hence we assume that λ

https://doi.org/10.1017/etds.2015.61 Published online by Cambridge University Press

and $\{\mu_\omega\}$ are given. However, in order to control the regularity of $\{h_\omega\}$, we employ a very similar construction to the one in [**6**].

Step 1: Construction of h_ω. For $\omega \in \Omega$ and $k \in \mathbb{N}$, let $f_{\omega,k} := \Lambda_k(\theta^{-k}\omega)^{-1}\mathscr{L}^k_{\theta^{-k}\omega}(1)$. By the same argument as in the proof of Proposition 2.1, we have, for $x, y \in [a]_\omega$ for some $a \in \mathscr{W}^1_\omega$, that

$$|f_{\omega,k}(x) - f_{\omega,k}(y)| \le \frac{\mathscr{L}^k_{\theta^{-k}\omega}(1)(y)(B_\omega - 1)d_r(x, y)}{\Lambda_k(\theta^{-k}\omega)} = f_{\omega,k}(y)(B_\omega - 1)d_r(x, y).$$

It follows from this and $\int \log B_\omega \, dP < \infty$ that $\log f_{\omega,k}$ is locally Hölder continuous with index 1 with Hölder constant bounded by $B_\omega - 1$. Moreover, by [**18**, Remark 4.2], the measure $\{\mu_\omega\}$ satisfies the following Gibbs property: for a.e. $\omega \in \Omega_{\mathrm{bi}}$, $k \in \mathbb{N}$ and $a \in \mathscr{W}^k_{\theta^{-k}\omega}$ and $x \in [a]_{\theta^{-k}\omega}$,

$$(\Lambda_k(\theta^{-k}\omega))^{-1} e^{S_k\varphi_{\theta^{-k}\omega}(x)} \le B_\omega(E_\omega)^{-1}\mu_{\theta^{-k}\omega}([a]_{\theta^{-k}\omega}), \tag{14}$$

where $E_\omega := \min_{b\in\mathscr{I}} \mu_\omega(T_{\theta^{-1}\omega}([b]_{\theta^{-1}\omega}))$ is strictly positive by finiteness of $\mathscr{I}$. By summing over all cylinders, we obtain that $f_{\omega,k}(x) \le B_\omega(E_\omega)^{-1}$ for a.e. $\omega \in \Omega_{\mathrm{bi}}$ and $k \in \mathbb{N}$. Moreover, [**18**, Corollary 4.1] shows that there exists $\Omega^* \subset \Omega$ of positive measure such that $f_{\omega,k}(x)$ is bounded from below whenever $\theta^{-k}\omega \in \Omega^*$. Hence, the random function (h_ω) defined by

$$h_\omega(x) := \liminf_{k\to\infty, \theta^{-k}\omega\in\Omega^*} f_{\omega,k}(x)$$

is bounded from above and below.

Step 2: Properties of h_ω. Observe that Fatou's lemma implies that $\mathscr{L}_\omega(h_\omega) \le \lambda_\omega h_{\theta\omega}$. Hence, for $\omega \in \Omega$ and $n \in \mathbb{N}$ with $\theta^n\omega \in \Omega_{\mathrm{bi}}$, we have

$$\int h_\omega \, d\mu_\omega = \int \Lambda_n(\omega)^{-1}\mathscr{L}^n_\omega(h_\omega)\, d\mu_{\theta^n\omega} \le \int h_{\theta^n\omega}\, d\mu_{\theta^n\omega} \le B_{\theta^n\omega}(E_{\theta^n\omega})^{-1}.$$

By ergodicity of θ, it follows that $\int h_\omega \, d\mu_\omega \le \operatorname{ess\text{-}inf}_{\omega\in\Omega_{\mathrm{bi}}} B_\omega(E_\omega)^{-1}$. By the same arguments as in [**6**], it then follows again from ergodicity that $\mathscr{L}_\omega(f_\omega) = \lambda_\omega f_{\theta\omega}$. It also follows from the above estimates that $\{\log h_\omega\}$ is locally Hölder continuous with index 1 and Hölder constant $B_\omega - 1$, and that $\|h_\omega\|_\infty \le B_\omega(E_\omega)^{-1}$ for a.e. $\omega \in \Omega_{\mathrm{bi}}$. In order to show that h_ω is bounded from below, note that finiteness of $\mathscr{I}$ and local Hölder continuity imply that, for a.e. $\omega \in \Omega_{\mathrm{bi}}$ and $x \in X_\omega$,

$$\lambda_{\theta^{-1}\omega}h_\omega(x) = \mathscr{L}_{\theta^{-1}\omega}(h_{\theta^{-1}\omega})(x) \ge \min_{a\in\mathscr{I}}\left(\inf_{y\in[a]_{\theta^{-1}\omega}} e^{\varphi_{\theta^{-1}\omega}(y)}h_{\theta^{-1}\omega}(y)\right) > 0.$$

Hence, $\inf h_\omega > 0$ for a.e. $\omega \in \Omega_{\mathrm{bi}}$. Combining these upper and lower bounds with positivity of $\mathscr{L}_\omega$ gives that $0 < \inf h_\omega \le \sup h_\omega < \infty$ for a.e. $\omega \in \Omega$. This proves (1) and (2) of Theorem 4.1.

Step 3: Normalizing the operator. Step 2 implies that $\log h_\omega(x)$ is locally Hölder with index 1 and, in particular, that $\log h_{\theta\omega}(T_\omega(x))$ is locally Hölder with index 2. Moreover, since $h_\omega(x)$ is bounded from above and below, $V_1^\omega(\log h_{\theta\omega} \circ T_\omega) < \infty$ a.s. In particular, the normalized potential defined by

$$\tilde{\varphi}_\omega(x) := \varphi_\omega(x) + \log h_\omega(x) - \log h_{\theta\omega}(T_\omega(x)) - \log \lambda_\omega$$

https://doi.org/10.1017/etds.2015.61 Published online by Cambridge University Press

satisfies property ($\hat{\mathrm{H}}$). Furthermore, we have

$$\Lambda_n(\omega)h_{\theta^n\omega}\tilde{\mathscr{L}}_\omega^n(f) = \mathscr{L}_\omega^n(f \cdot h_\omega) \tag{15}$$

for a.e. ω, all $n \in \mathbb{N}$ and each bounded and continuous function $f : X_\omega \to \mathbb{R}$. In particular, $\tilde{\mathscr{L}}_\omega(1) = 1$ and, for $\{\nu_\omega\}$ defined by $d\nu_\omega := h_\omega\, d\mu_\omega$, we have $\tilde{\mathscr{L}}_\omega^*(\nu_{\theta\omega}) = \nu_\omega$. Now let $t \in (0, 1)$, $c > 0$ and $(l_n(\omega))$ be given by Corollary 3.1. For a fibrewise Lipschitz function f with $D(f) \le 1$, $n \in \mathbb{N}$ and $x \in X_{\theta^{l_n}(\omega)}$, the Monge–Kantorovich duality and (1) in Corollary 3.1 then imply that

$$\begin{aligned}\left|\tilde{\mathscr{L}}_\omega^{l_n}(f)(x) - \int f\, d\nu_\omega\right| &= \left|\int f\, d(\tilde{\mathscr{L}}_\omega^{l_n})^*(\delta_x) - \int f\, d(\tilde{\mathscr{L}}_\omega^{l_n})^*(\nu_{\theta^{l_n}\omega})\right| \\ &\le W((\tilde{\mathscr{L}}_\omega^{l_n})^*(\delta_x), (\tilde{\mathscr{L}}_\omega^{l_n})^*(\nu_{\theta^{l_n}\omega})) \le ct^n W(\delta_x, \nu_\omega) \\ &= c \cdot t^n \int d(x, y)\, d\nu_\omega(y) \le ct^n. \end{aligned} \tag{16}$$

Hence, $\|\tilde{\mathscr{L}}_\omega^{l_n}(f) - \int f\, d\nu_\omega\|_\infty \le ct^n$. Combining this with (2) of Corollary 3.1 then proves the first inequality in Theorem 4.2, and the second follows by the same arguments. Finally, the uniqueness of $\{\nu_\omega\}$ clearly follows from (1) in Corollary 3.1. This proves Theorem 4.2.

Step 4: Transferring the results. By (15), we have

$$\frac{\mathscr{L}_\omega^n(f)}{\Lambda_n(\omega)h_{\theta^n(\omega)}} - \int f\, d\mu_\omega = \tilde{\mathscr{L}}_\omega^n(f/h) - \int (f/h)\, d\nu_\omega. \tag{17}$$

Hence, parts (6) and (7) in Theorem 4.1 can be proved through estimating $D_\omega(f/h)$. By dividing the supremum below into x, y according to $d(x, y) = 1$ or $d(x, y) < 1$ and then applying locally Hölder continuity with index 1 of $\log h$ in the second case, it follows that

$$\begin{aligned} D_\omega(f/h) &\le \|1/h_\omega\|_\infty D_\omega(f) + \|f_\omega\|_\infty D(1/h_\omega) \\ &\le \|1/h_\omega\|_\infty \left(D_\omega(f) + \|f_\omega\|_\infty \sup_{x,y \in X_\omega} \frac{|h_\omega(x)/h_\omega(y) - 1|}{d(x, y)} \right) \\ &\le \|f_\omega\|_L \left\| \frac{1}{h_\omega} \right\|_\infty \max\left\{ 1, \|h_\omega\|_\infty \left\| \frac{1}{h_\omega} \right\|_\infty - 1, B_\omega - 1 \right\} \\ &=: \|f_\omega\|_L K_\omega/2. \end{aligned} \tag{18}$$

Combining the estimate with (17) and Theorem 4.2 proves (6) of Theorem 4.1. We will proceed with the proof of (7). In order to do so, note that Corollary 3.2 implies, using (16), that

$$\left|\tilde{\mathscr{L}}_\omega^n(f)(x) - \int f\, d\nu_\omega\right| \le c_\omega^* s^n W(\delta_x, \nu_{\theta^n\omega}) \le c_\omega^* s^n,$$

$$\left|\tilde{\mathscr{L}}_\omega^n(f)(x) - \tilde{\mathscr{L}}_\omega^n(f)(y)\right| \le c_\omega^* s^n W(\delta_x, \delta_y) = c_\omega^* s^n d(x, y),$$

where f satisfies $D(f) \le 1$. Combining (17) with the estimate on $D(f/h)$ proves (7).

Step 5: Uniqueness. Hence, in order to prove Theorem 4.1, it remains to deduce the uniqueness of $\{h_\omega\}$ and $\{\mu_\omega\}$. So assume that $\{\mu_\omega^{(i)}\}$ with $i = 1, 2$ are random probability measures with $\mathscr{L}_\omega^*(\mu_{\theta\omega}^{(i)}) = \lambda(\omega)\mu_\omega^{(i)}$. Then $\{h_\omega\, d\mu_\omega^{(i)}\}$ are both $\tilde{\mathscr{L}}_\omega$-invariant and hence equal by uniqueness from Theorem 4.2. The uniqueness of $\{h_\omega\, d\mu_\omega^{(i)}\}$ follows from the same arguments. Hence, also Theorem 4.1 is proved. □

https://doi.org/10.1017/etds.2015.61 Published online by Cambridge University Press

Acknowledgements. The author would like to thank the anonymous referee for helpful suggestions and pointing out the applicability to non-stationary shift spaces, which improved the content and exposition of this note. Furthermore, the author would like to express his deep gratitude to the Erwin–Schrödinger Institute in Vienna for warm hospitality and excellent working conditions and the DFG-network 'Skew product dynamics and multifractal analysis' for financial support: the ideas of proofs emerged while preparing and presenting a mini course at the Workshop on complexity and dimension theory of skew product systems. The author also acknowledges support by Fundação para Ciência e a Tecnologia through project PTDC/MAT/120346/2010.

References

[**1**] P. Arnoux and A. M. Fisher. Anosov families, renormalization and non-stationary subshifts. *Ergod. Th. & Dynam. Sys.* **25** (2005), 661–709.
[**2**] T. Bogenschütz and V. M. Gundlach. Ruelle's transfer operator for random subshifts of finite type. *Ergod. Th. & Dynam. Sys.* **15** (1995), 413–447.
[**3**] X. Bressaud, R. Fernández and A. Galves. Decay of correlations for non-Hölderian dynamics: A coupling approach. *Electron. J. Probab.* **4** (1999), 19 (electronic).
[**4**] J. Buzzi. Exponential decay of correlations for random Lasota–Yorke maps. *Comm. Math. Phys.* **208** (1999), 25–54.
[**5**] H. Crauel. *Random Probability Measures on Polish Spaces*. Taylor & Francis, London, 2002.
[**6**] M. Denker, Y. Kifer and M. Stadlbauer. Thermodynamic formalism for random countable Markov shifts. *Discrete Contin. Dyn. Syst.* **22** (2008), 131–164.
[**7**] W. Doeblin and R. Fortet. Sur des chaînes à liaisons complètes. *Bull. Soc. Math. France* **65** (1937), 132–148.
[**8**] I. V. Evstigneev and S. A. Pirogov. Stochastic nonlinear Perron–Frobenius theorem. *Positivity* **14** (2010), 43–57.
[**9**] A. M. Fisher. Nonstationary mixing and the unique ergodicity of adic transformations. *Stoch. Dyn.* **9** (2009), 335–391.
[**10**] S. Galatolo and M. J. Pacifico. Lorenz-like flows: Exponential decay of correlations for the Poincaré map, logarithm law, quantitative recurrence. *Ergod. Th. & Dynam. Sys.* **30** (2010), 1703–1737.
[**11**] M. Hairer and J. C. Mattingly. Spectral gaps in Wasserstein distances and the 2D stochastic Navier–Stokes equations. *Ann. Probab.* **36** (2008), 2050–2091.
[**12**] H. Hennion. Sur un théorème spectral et son application aux noyaux lipchitziens. *Proc. Amer. Math. Soc.* **118** (1993), 627–634.
[**13**] Y. Kifer. Thermodynamic formalism for random transformations revisited. *Stoch. Dyn.* **8** (2008), 77–102.
[**14**] Y. Kifer and P.-D. Liu. Random dynamics. *Handbook of Dynamical Systems.* Vol. 1B. Elsevier BV, Amsterdam, 2006.
[**15**] S. Orey. Markov chains with stochastically stationary transition probabilities. *Ann. Probab.* **19** (1991), 907–928.
[**16**] O. M. Sarig. Existence of Gibbs measures for countable Markov shifts. *Proc. Amer. Math. Soc.* **131** (2003), 1751–1758.
[**17**] W. Shen and S. van Strien. On stochastic stability of expanding circle maps with neutral fixed points. *Dyn. Syst.* **28** (2013), 423–452.
[**18**] M. Stadlbauer. On random topological Markov chains with big images and preimages. *Stoch. Dyn.* **10** (2010), 77–95.
[**19**] M. Stadlbauer. Corrigendum and addendum to: On random topological Markov chains with big images and preimages. *Stoch. Dyn.* **14** (2014), 1492001, 3 pp.
[**20**] C. T. I. Tulcea and G. Marinescu. Théorie ergodique pour des classes d'opérations non complètement continues. *Ann. of Math. (2)* **52** (1950), 140–147.
[**21**] C. Villani. *Optimal Transport*. Springer, Berlin, 2009.
[**22**] X. Zhang. Stochastic Monge–Kantorovich problem and its duality. *Stochastics* **85** (2013), 71–84.

https://doi.org/10.1017/etds.2015.61 Published online by Cambridge University Press