A BACKWARD ERGODIC THEOREM
AND ITS FORWARD IMPLICATIONS

ANUSH TSERUNYAN AND JENNA ZOMBACK

Abstract. We prove a backward pointwise ergodic theorem for countable-to-one probability-measure-preserving (pmp) transformations, where the ergodic averages are taken along trees in the backward orbit of the point. This yields a new (forward) pointwise ergodic theorem along trees for pmp actions of free groups, strengthening an earlier (from 2000) theorem of Bufetov. We also discuss other applications of the backward theorem, in particular to the shift map on Markov chains, which yields a (forward) pointwise ergodic theorem along trees for the natural action of the free groups on their boundary.

1. Introduction

The classical pointwise ergodic theorem, whose first instance dates back to Birkhoff [Bir31], states that a measure-preserving (i.e., $T_*\mu = \mu$) transformation $T : X \to X$ on a standard probability space $(X,\mu)$ is ergodic (i.e. every $T$-invariant measurable set is either null or conull) if and only if for every $f \in L^1(X,\mu)$, for a.e. $x \in X$, \[ \lim_{n \to \infty} \frac{1}{n} \sum_{k=0}^{n-1} f(T^k x) = \int f \, d\mu. \]

More generally, pointwise ergodic theorems have been proven for probability-measure-preserving (pmp) actions of finitely generated (f.g.) semigroups, but with weighted ergodic averages. One of the most general results in this vein is [Buf00, Theorem 1], which applies to pmp actions of semigroups generated by a finite set $I$. It states the convergence of weighted ergodic averages taken over the balls $I \leq n$. The weights are assigned to the finite words in $I$ by a Markov chain on $I$. We state the version of this theorem for free groups as Theorem 1.6.

1.1. Results. We prove the following new pointwise ergodic theorem for pmp actions of free groups, strengthening the version of Bufetov’s theorem for free groups (Theorem 1.6): we replace balls by arbitrary trees. See Fig. 1 for an example of such a tree.

**Theorem 1.1** (Pointwise ergodic along trees). Let $F_r$ be the free group on $r$ generators, $r < \infty$, and let $I$ be the standard symmetric set of generators. Let $F_r \curvearrowright (X,\mu)$ be a (not necessarily free) pmp action of $F_r$. Let $P$ be a Markov chain on $I$ given by a stationary distribution $\pi$ and transition matrix $P$, where for all $a, b \in I$, $\pi(a) > 0$, and $P(a,b) = 0$ iff $a = b^{-1}$; this induces a probability distribution on $I^n$ for each $n \in \mathbb{N}$, which we still denote by $P$. Then for any $f \in L^1(X,\mu)$, for $\mu$-a.e. $x \in X$, \[ \frac{1}{P(S)} \sum_{w \in S} f(w \cdot x)P(w) \to \overline{f}(x) \text{ as } P(S) \to \infty \]

where $S \subseteq I^{<\mathbb{N}}$ ranges over all finite right-rooted trees (see Fig. 1) on the alphabet $I$ and $\overline{f}$ is the conditional expectation of $f$ with respect to the $\sigma$-algebra of $F_r$-invariant Borel sets.

Date: December 22, 2020.

The authors’ research was partially supported by NSF Grant DMS-1855648 and the first author was also partially supported by NSERC Discover Grant RGPIN-2020-07120.

\(^1\)In ergodic theory, a set $A \subseteq X$ is said to be $T$-invariant if $T^{-1}(A) = A$. 
This theorem is stated more precisely as Theorem 6.1 and the definition of right-rooted trees is given in Section 4.1.

We also prove a pointwise ergodic theorem for the (necessarily non-pmp) action of the free group \( \mathbb{F}_r \) on \( r \leq \infty \) generators on its boundary \( \partial \mathbb{F}_r \) (here, we identify \( \partial \mathbb{F}_r \) with the space of infinite reduced words on the standard symmetric set of generators of \( \mathbb{F}_r \)).

**Theorem 1.2** (Pointwise ergodic for \( \mathbb{F}_r \) \( \curvearrowright \) \( \partial \mathbb{F}_r \)). Let \( I \) be the standard symmetric set of generators of \( \mathbb{F}_r \) \( (r \leq \infty) \), and \( \mathbb{F}_r \) \( \curvearrowright \) \( \partial \mathbb{F}_r \) be the natural action of \( \mathbb{F}_r \) on its boundary \( \partial \mathbb{F}_r \subseteq I^\mathbb{N} \). Let \( P \) be a Markov chain on \( I \) with stationary initial distribution and transition matrix \( P \) satisfying \( P(a,a^{-1}) = 0 \), for each \( a \in I \). Let \( E_T \) denote the orbit equivalence relation induced by the action and let \( \rho : E_T \rightarrow \mathbb{R}^+ \) be the cocycle induced by \( \rho(i \cdot x, x) := \frac{\pi(i)}{\pi(x_0)} P(i, x_0) \), for \( i \in I \). For each \( f \in L^1(\partial \mathbb{F}_r, P) \), for \( P \)-a.e. \( x \in \partial \mathbb{F}_r \), for all \( n \in \mathbb{N} \) and \( \emptyset \neq S \subseteq I^{\leq n} \),

\[
\frac{1}{|S \cdot x|_{\rho}} \sum_{w \in S} f(w \cdot x) \rho(w \cdot x, x) \rightarrow \tilde{f}(x) \text{ as } |S \cdot x|_{\rho} \rightarrow \infty
\]

where \( S \subseteq I^{<\mathbb{N}} \) ranges over all right-rooted trees (see Section 4.1) of reduced words on \( I \) of bounded length, \( |S \cdot x|_{\rho} := \sum_{w \in S} \rho(w \cdot x, x) \), and \( \tilde{f} \) is the conditional expectation of \( f \) with respect to the \( \sigma \)-algebra of \( \mathbb{F}_r \)-invariant Borel sets.

A more precise version of this is stated as Theorem 5.13.

Note that even though \( \mathbb{F}_r \) is nonamenable (for \( 2 \leq r \leq \infty \)) and its boundary action is free on a cocountable set, its orbit equivalence relation is (Borel) hyperfinite (hence, amenable, see [Kec20, 8.2 and 8.3]) by [DJK94, Corollary 8.2] because it is induced by the shift map. Thus, the action does not admit an invariant probability measure [JKL02, Proposition 1.7].

We also provide an example of an ergodic measure on \( \partial \mathbb{F}_r \) for each \( 2 \leq r \leq \infty \) for which Theorem 5.13 holds. In particular, we define a shift-invariant ergodic measure on \( \partial \mathbb{F}_\infty \). The authors are not aware of any such measure being exhibited explicitly in the literature.

The main result underlying Theorem 1.1 and Theorem 1.2 is a backward (i.e., in the direction of \( T^{-1} \)) pointwise ergodic theorem for probability-measure-preserving (pmp) Borel transformations \( T \). Although \( T \) is pmp, the induced orbit equivalence relation \( E_T \) is not pmp, unless \( T \) is one-to-one, so the averages are weighted according to the corresponding Radon–Nikodym cocycle \( \rho : E_T \rightarrow \mathbb{R}^+ \). This theorem states convergence of ergodic averages.
along trees of preimages under $T$. We first state the following already interesting special case, which is less technical, and, in the case where $f$ is bounded, can be deduced directly from the classical (forward) pointwise ergodic theorem. (The authors would be interested to know if it follows from the classical pointwise ergodic theorem for an unbounded $f$.) This is later restated as Corollary 4.11.

**Corollary 1.3** (Backward pointwise ergodic along complete trees). Let $(X, \mu)$ be a standard probability space and $T : X \to X$ an aperiodic\(^2\) countable-to-one pmp Borel transformation. Let $E_T$ denote the induced orbit equivalence relation and $\rho : E_T \to \mathbb{R}^+$ the Radon–Nikodym cocycle corresponding to $\mu$. Then for any $f \in L^1(X, \mu)$, for $\mu$-a.e. $x \in X$, for all $i \in \mathbb{N}$, $\sum_{y \in T^{-i}(x)} |f(y)|\rho(y, x) < \infty$, and

$$
\lim_{n \to \infty} \frac{1}{n+1} \sum_{i=0}^{n} \sum_{y \in T^{-i}(x)} f(y)\rho(y, x) = \overline{f}(x),
$$

where $\overline{f}$ is the conditional expectation of $f$ with respect to the $\sigma$-algebra of $T$-invariant Borel sets. Furthermore, if $f \in L^p(X, \mu)$, $1 \leq p < \infty$, then $\overline{f} \in L^p$, and

$$
\frac{1}{n+1} \sum_{i=0}^{n} \sum_{y \in T^{-i}(x)} f(y)\rho(y, x) \to_{L_p} \overline{f}(x) \text{ as } n \to \infty.
$$

Thus, while the classical pointwise ergodic theorem for an ergodic $T$ allows us to start almost anywhere in the space and walk forward along the graph of $T$ to calculate the integral of $f$, Corollary 1.3 allows us to walk backward in all directions along the graph of $T$ to calculate this integral. Note that this implies the usual (forward) pointwise ergodic theorem for one-to-one transformations $T$ when Corollary 1.3 is applied to $T^{-1}$.

Corollary 1.3 can be viewed as an ergodic theorem for the Ruelle transfer operator [FJ01a, FJ01b, Rue02] $\mathcal{L}$ acting on $L^1(X, \mu)$ by

$$(\mathcal{L}f)(x) := \sum_{y \in T^{-1}(x)} f(y)g(y),$$

for the special case where $T : X \to X$ is a countable-to-one pmp transformation and $g(y) := \rho(y, Ty)$, and where $\rho : E_T \to \mathbb{R}^+$ is the Radon–Nikodym cocycle corresponding to $\mu$. Thus, $(\mathcal{L}^n f)(x)$ is the $\rho$-weighted average of $f$ on $T^{-n}(x)$, and Corollary 1.3 says that for any $f \in L^p$, $\frac{1}{n+1} \sum_{k \leq n} (\mathcal{L}^n f)(x)$ converges to $\overline{f}(x)$ both $\mu$-a.e. and in $L^p$.

We now state the more general version of Corollary 1.3, which implies all the results mentioned above. It is restated later as Theorem 4.7.

**Theorem 1.4** (Backward pointwise ergodic along trees). Let $(X, \mu)$ be a standard probability space and $T : X \to X$ an aperiodic countable-to-one pmp Borel transformation. Let $E_T$ denote the induced orbit equivalence relation and $\rho : E_T \to \mathbb{R}^+$ the Radon–Nikodym cocycle corresponding to $\mu$. Then for any $f \in L^1(X, \mu)$, for $\mu$-a.e. $x \in X$, for all $n \in \mathbb{N}$ and $S_x \subseteq \bigcup_{i=0}^{n} T^{-i}(x)$, $\sum_{y \in S_x} |f(y)|\rho(y, x) < \infty$, and

$$
\frac{1}{|S_x|^x} \sum_{y \in S_x} f(y)\rho(y, x) \to \overline{f}(x) \text{ as } |S_x|^x \rho \to \infty,
$$

\(^2\)T is called aperiodic if for all $x \in X$ and $n \in \mathbb{N} \setminus \{0\}$, $T^n(x) \neq x.$
where $S_x$ ranges over all (possibly infinite) subtrees of the graph of $T$ of finite height rooted at $x$ and directed towards $x$, $|S_x|^p := \sum_{y \in S_x} \rho(y, x)$, and $\bar{f}$ is the conditional expectation of $f$ with respect to the $\sigma$-algebra of $T$-invariant Borel sets.

We obtain Theorem 1.1 and Theorem 1.2 by applying Theorem 1.4 to specific choices of $T$. In addition to Theorem 1.1 and Theorem 1.2, we also explore other applications of Theorem 1.4 to the shift map on spaces of infinite words in Section 5. We recall the class of Markov chains on these spaces that are shift-invariant and shift-ergodic, and point out that for each such chain, Theorem 1.4 allows us to calculate the expectation of an $L^1$ function by walking backwards along trees.

1.2. Context and history. In general, a measure-preserving action of a countable (discrete) semigroup $G$ on a standard probability space $(X, \mu)$ is said to have the pointwise ergodic property along a sequence $(F_n)$ of finite subsets of $G$, if for every $f \in L^1(X, \mu)$, for $\mu$-a.e. $x \in X$,

$$\lim_{n \to \infty} \frac{1}{|F_n|} \sum_{g \in F_n} f(g \cdot x) = \bar{f},$$

where $\bar{f}$ is the conditional expectation of $f$ with respect to the $\sigma$-algebra of $G$-invariant measurable sets, so in case the action is ergodic, the limit is just the expectation $\int_X f d\mu$ of $f$.

1.2.i. For pmp actions. It is a celebrated theorem of Lindenstrauss [Lin01] that the pointwise ergodic property is true for the pmp actions of all countable amenable groups along tempered Følner sequences and this was extended by Butkevich in [But00] to all countable left-cancellative amenable semigroups.

Amenability, or rather the fact that the Følner sets $F_n$ are almost invariant, is essential for the pointwise ergodic property as it ensures that the limit of averages is an invariant function. This is why, to obtain a version of the pointwise ergodic property for nonamenable (semi)groups, e.g. for the nonabelian free groups $F_r$, one has to imitate the almost invariance of finite sets by taking weighted averages instead, so that the weight of the boundary is small. The first instance of this was proven independently by Grigorchuk [Gri87,Gri99,Gri00] and by Nevo [Nev94]:

**Theorem 1.5** (Grigorchuk 1986; Nevo 1994). Let $F_r \curvearrowright (X, \mu)$ be a pmp action of the free group on $r < \infty$ generators on a standard probability space. For any $f \in L^1(X, \mu)$, for $\mu$-a.e. $x \in X$,

$$\lim_{n \to \infty} \frac{1}{n+1} \sum_{w \in B_n} f(w \cdot x)\mathbb{P}(w) = \bar{f}(x),$$

where $B_n$ is the (closed) ball of radius $n$ in the standard symmetric Cayley graph of $F_r$, $\mathbb{P}$ is the uniform probability distribution on each sphere of this graph$^3$, and $\bar{f}$ is the conditional expectation of $f$ with respect to the $\sigma$-algebra of $F_r$-invariant Borel sets.

This was later vastly generalized by Bufetov in [Buf00, Theorem 1] to a large class of stationary Markov chains (in lieu of the uniform distribution on spheres) and finitely generated semigroups. To stay within our context, we state his theorem here for free groups and a more special class of Markov chains.

---

$^3$This means that $\mathbb{P}(w) := \frac{1}{2r(2r-1)^{|w|-1}}$ for $w \neq 1_r$, and $\mathbb{P}(1_r) := 1$. 

4
Theorem 1.6 (Bufetov 2000). Let \( \mathbb{F}_r \) be the free group on \( r \) generators, \( r < \infty \), and let \( I \) be the standard symmetric set of generators. Let \( \mathbb{F}_r \rtimes (X, \mu) \) be a (not necessarily free) pmp action of \( \mathbb{F}_r \). Let \( \mathbb{P} \) be a Markov chain on \( I \) given by a stationary distribution \( \pi \) and transition matrix \( P \), where for all \( a, b \in I \), \( \pi(a) > 0 \), and \( P(a, b) = 0 \) iff \( a = b^{-1} \); this induces a probability distribution on \( I^n \) for each \( n \in \mathbb{N} \), which we still denote by \( \mathbb{P} \). Then for any \( f \in L^1(X, \mu) \), for \( \mu \)-a.e. \( x \in X \),

\[
\lim_{n \to \infty} \frac{1}{n+1} \sum_{w \in I \leq n} f(w \cdot x) \mathbb{P}(w) = \overline{f}(x),
\]

where \( \overline{f} \) is the conditional expectation of \( f \) with respect to the \( \sigma \)-algebra of \( \mathbb{F}_r \)-invariant Borel sets.

Indeed, Theorem 1.5 is a special case of this: let \( I \) be the standard symmetric set of generators of \( \mathbb{F}_r \), and let \( \mathbb{P} \) be the Markov chain with the uniform initial distribution on \( I \) (i.e. the constant \( \frac{1}{2r} \) vector) and transition probabilities \( P(a, a^{-1}) = P(a^{-1}, a) = 0 \) for all \( a \in I \), and all other transition probabilities equal (to \( \frac{1}{2r-1} \)).

Theorem 1.6, in turn, is a special case of our Theorem 1.1, taking the limit along the sequence \( S_n := I \leq n = \bigcup_{i \leq n} I^i \) and observing that \( \mathbb{P}(S_n) = n + 1 \).

Finally, Theorem 1.1 is deduced from Theorem 1.4 by applying the latter to an auxiliary transformation on \( X \times \partial \mathbb{F}_r \).

Other instances of pointwise ergodic theorems are known for pmp actions of finitely generated groups where the averages are taken over sets that are not trees. For example, [NS94] includes that the pointwise theorem is true with ergodic averages taken over spheres (in case of the free group), and more generally, in [Buf02], they are taken over the sets \( I^n \) (in our notation). More general treatment of pointwise ergodic theorems for groups is given by Bowen and Nevo in [BN13] and [BN15, Theorems 6.2 and 6.3]. We refer to [BK12] for a survey of pointwise ergodic theorems for groups.

1.2.ii. For nonsingular (null-preserving) actions. There is a suitable analogue of the pointwise ergodic property for merely nonsingular actions of (semi)groups on a standard probability space: the averages have to be weighted by the corresponding Radon–Nikodym cocycle [KM04, Section 8]. Much less is known for such actions: a pointwise ergodic theorem for \( \mathbb{Z}^d \) was first proven by Feldman [Fel07] and then generalized in two different directions by Hochman [Hoc10] and by Dooley and Jarrett [DJ16]. For general groups of polynomial growth, Hochman obtained [Hoc13, Theorem 1.4] a slightly weaker form of a nonsingular ergodic theorem where the a.e. convergence is replaced with a.e. convergence in density.

On the negative side, M. Hochman proved [Hoc13, Theorem 1.1] that the pointwise ergodic theorem for null-preserving actions holds only along sequences of subsets of the group satisfying the so-called Besicovitch covering property. He then infers [Hoc13, Theorems 1.2 and 1.3] that the null-preserving pointwise ergodic theorem fails for any sequence of subsets of \( \bigoplus_{n \in \mathbb{N}} \mathbb{Z} \) and any subsequence of balls in nonabelian free groups as well as in the Heisenberg group.

---

4A measurable action of a countable semigroup \( G \) on a probability space \((X, \mu)\) is called nonsingular (or null-preserving) if for each \( g \in G \), \( g_* \mu \sim \mu \), equivalently, the \( g \)-preimage of a null set is null.
1.2.iii. Hybrid setting. Our main result (Theorem 1.4) is a contribution to the ergodic theory of actions of finitely generated (semi)groups in both the pmp and null-preserving settings. As mentioned above, although Theorem 1.4 is about a pmp transformation $T$, its orbit equivalence relation $E_T$ is generally not pmp, and assuming, as we may, that $E_T$ is null-preserving, the ergodic averages are weighted with the corresponding Radon–Nikodym cocycle. Furthermore, Theorem 1.4 implies a forward ergodic theorem for the natural action of the free group $F_r$ on its boundary (Theorem 1.2), which is merely null-preserving. On the other hand, Theorem 1.4 implies Theorem 1.1, which applies to pmp actions.

1.3. A word on the proof of the backward ergodic theorem (Theorem 1.4). The pointwise ergodic property equates the global condition of ergodicity with the local (pointwise) combinatorics of the action. Most known proofs of instances of the pointwise ergodic property implicitly or explicitly have a combinatorial part to them, manifested as a tiling or covering argument. In particular, the proofs by Keane and Petersen in [KP06] and Tserunyan in [Tse18] of the classical pointwise ergodic theorem explicitly feature a tiling argument. Tserunyan later formulated a tiling property that implies a corresponding pointwise ergodic theorem for any countable group along a sequence of finite sets, which was explicitly stated and used in [BZ19]. This tiling property is what allows us to deal with unbounded functions because one can tile arbitrarily well after restricting the function to a very large set on which it is bounded.

Here, we prove a tiling property (Lemma 4.1) for countable-to-one pmp Borel transformations $T$, where the points are relatively weighted by the corresponding Radon–Nikodym cocycle $\rho$ on the induced orbit equivalence relation $E_T$. We tile sets of the form $\nabla^y_x \cdot x := \bigcup_{i \leq n} T^{-i}(x)$. Our tiles are of the form $S_y$, where $S_y$ is an arbitrary subtree of the graph of $T$ of finite height rooted at $y$ and directed towards $y$, and $y \in \bigcup_{i \leq n} T^{-i}(x)$ (see Fig. 2 for when $T$ is the shift function $s$ on $2^\mathbb{N}$). Then Theorem 1.4 follows from this property and the $T$-invariance of the limit in Theorem 1.4, with the use of the Luzin–Novikov uniformization theorem [Kec95, Theorem 18.10].

Like the classical ergodic theorem, it is not clear a priori that the limit (or rather, the lim sup) in Theorem 1.4, and especially in its special case Corollary 4.11, is invariant, and the

---

**Figure 2.** $\nabla^4_s(x)$ with examples of $S_x$ and $S_{1-x}$ circled
invariance of the lim sup is crucial for our deduction of the ergodic theorem from the tiling property. The issue is that even for the special case of complete trees, i.e. sets of the form $\triangledown^n_T \cdot x := \bigcup_{i=0}^n T^{-i}(x)$, the averages of a function $f \in L^1(X, \mu)$ over $\triangledown^n_T \cdot x$ and over $\triangledown^m_T \cdot T(x)$ may defer by an uncontrollable amount for any $n, m \in \mathbb{N}$ (even for $m = n + 1$ or $n$) because the $\rho$-weight of $(\triangledown^n_T \cdot x) \triangle (\triangledown^m_T \cdot T(x))$ grows with $n, m$. This is what leads to using arbitrary trees, but even with them, comparing the averages over the trees $S_x$ and $S_x \cup \{T(x)\}$ only gives that the lim sup is nondecreasing in the direction of $T$, so one has to apply Poincaré recurrence (or equivalently, nonsmoothness of $E_T$) to deduce that the lim sup is constant on the orbit of $x$.

Organization. In Section 2, we give the necessary definitions and notation that are used throughout the paper, and we prove some preliminary lemmas about countable-to-one pmp Borel transformations $T$. In Section 3, we use the fact that $T$ is pmp to show that the $\rho$-weight of each column in $\triangledown^n_T \cdot x$ is equal to the $\rho$-weight of the point $x$ (Lemma 3.6), which allows us to compute integrals by averaging over the tiles (Corollary 3.7). In Section 4, we explicitly state and prove the suitable tiling property (Lemma 4.1) and deduce Theorem 1.4 from it. In Sections 5 and 6, we provide examples of countable-to-one pmp ergodic Borel transformations (to which Theorem 1.4 applies), and deduce the corresponding ergodic theorems, in particular, for pmp actions of free groups and the boundary action of the free groups.

Acknowledgements. The authors owe many thanks to Gabor Elek for encouraging them to investigate the pointwise ergodic property for the boundary action of the free groups. Also, thanks to Florin Boca and Grigory Terlov for helpful discussions, and to Alexander Kechris and Aristotelis Panagiotopoulos for inviting the second author to speak in the Caltech seminar, which sparked many of the questions addressed in the paper. We thank Anton Bernshteyn for pointing out that Corollary 1.3 for bounded functions follows from the classical pointwise ergodic theorem (for one pmp transformation). We are also grateful to Ronnie Chen for inspiring discussion and for many useful comments and corrections. Finally, the authors owe a huge acknowledgement to Lewis Bowen for many useful suggestions and insight, and in particular for pointing out Bufetov’s theorem [Buf00], which led the authors to the results in Section 6.

2. Preliminaries

Throughout, let $(X, \mu)$ be a standard probability space. A countable Borel equivalence relation on $(X, \mu)$ is an equivalence relation that is a Borel subset of $X^2$ whose every equivalence class is countable. By the Luzin–Novikov uniformization theorem [Kec95, Theorem 18.10], if $B \subseteq X$ is Borel, then so is $[B]_E := \bigcup_{x \in B} [x]_E$.

We say that a countable Borel equivalence relation $E$ on $(X, \mu)$ is $\mu$-preserving (resp. null-preserving) if for any partial Borel injection $\gamma : X \to X$ with $\text{graph}(\gamma) \subseteq E$, $\mu(\text{dom}(\gamma)) = \mu(\text{im}(\gamma))$ (resp. $\text{dom}(\gamma)$ is null if and only if $\text{im}(\gamma)$ is null).

Note that $E$ is null-preserving if and only if the $E$-saturations of null sets are null. By [KM04, Section 8], a null-preserving $E$ admits a so-called Radon–Nikodym cocycle $\rho : E \to \mathbb{R}^+$ corresponding to $\mu$. Being a cocycle for a function $\rho : E \to \mathbb{R}^+$ means that it satisfies the cocycle identity

$$\rho(x, y)\rho(y, z) = \rho(x, z),$$

for all $E$-equivalent $x, y, z \in X$. We say that $\rho$ is the Radon–Nikodym cocycle corresponding to $\mu$ (or that $\mu$ is $\rho$-invariant) if it is Borel (as a real-valued function on the standard Borel
space $E$) and for any partial Borel injection $\gamma : X \to X$ with graph($\gamma$) $\subseteq E$ and $f \in L^1(X, \mu)$,
\[
\int_{\text{im}(\gamma)} f(x) \, d\mu(x) = \int_{\text{dom}(\gamma)} \rho(\gamma(x), x) f(\gamma(x)) \, d\mu(x).
\] (2.1)

We call a subset $S$ of an $E$-equivalence class $C$ $\rho$-finite if $|S|^\rho = \sum_{y \in S} \rho(y, x) < \infty$ for some $x \in C$ (although the value $|S|^\rho$ depends on the choice of $x$, its finiteness does not, by the cocycle identity). Further, for a non-negative function $f : X \to \mathbb{R}$, we define the $\rho$-weighted average of $f$ over $S$ by
\[
A^\rho_f[S] := \frac{\sum_{y \in S} f(y) \rho(y, x)}{|S|^\rho}
\]
for some $x \in C$. Again, this value does not depend on the choice of $x \in C$ by the cocycle identity. We also use the same notation for a general real-valued function $f$, provided $A^\rho_f[S] < \infty$.

We say that a Borel transformation $T : X \to X$ is $\mu$-preserving (resp. null-preserving) if $T_*\mu = \mu$ (resp. $T_*\mu \sim \mu$). Let $E_T$ denote the induced equivalence relation on $X$, that is:
\[
x E_T y :\Leftrightarrow \exists n, m, T^n(x) = T^m(y).
\]
Note that if $T$ is countable-to-one, $E_T$ is countable (i.e. each $E$-class is countable).

Even when $T$ is $\mu$-preserving, $E_T$ may not be $\mu$-preserving since $T$ may not be injective. In fact, $E_T$ may not even be null-preserving (it is possible for the $T$-image of a null set to have positive measure). However, the following result (originally proven by Kechris using Woodin’s argument for the analogous statement for Baire category) shows that we may neglect this issue. Three different proofs of this are given in [Mil04, Proposition 2.1] and [Mil17, 1.3], and we give a fourth one here, which is a measure-exhaustion argument.

**Lemma 2.2** (Kechris–Woodin). Let $E$ be a countable Borel equivalence relation on a standard probability space $(X, \mu)$. Then $E$ is null-preserving when restricted to some conull set.

**Proof.** Using the Feldman–Moore theorem [FM77], fix a countable set $\Gamma$ of Borel involutions $\gamma : X \to X$ such that $E = \bigcup_{\gamma \in \Gamma} \text{graph}(\gamma)$.

**Claim.** For a Borel set $Y \subseteq X$, and $\gamma \in \Gamma$, let $\gamma_Y := \gamma \cap Y^2$, i.e., the restriction of $\gamma$ to the set $\{y \in Y \cap \text{dom}(\gamma) : \gamma(y) \in Y\}$. Then $E|_Y$ is null-preserving if for each $\gamma \in \Gamma$ and each Borel $B \subseteq \text{dom}(\gamma_Y)$, $\mu(B) > 0$ implies $\mu(\gamma_Y(B)) > 0$.

**Proof of Claim.** Let $B \subseteq Y$ be a set whose saturation $[B]_{E|_Y}$ has positive measure. Then since $[B]_{E|_Y} = \bigcup_{\gamma \in \Gamma} \gamma_Y(B \cap \text{dom}(\gamma_Y))$, we must have $\mu(\gamma_Y(B \cap \text{dom}(\gamma_Y))) > 0$ for some $\gamma \in \Gamma$, which implies $\mu(B) \geq \mu(\gamma^2_Y(B \cap \text{dom}(\gamma_Y))) > 0$, by the assumption.

To construct a conull set $Y$ satisfying the hypothesis of the claim, it is enough to fix $\gamma \in \Gamma$ and find a conull set $Y_\gamma$ such that for each Borel $B \subseteq \text{dom}(\gamma_Y)$, $\mu(B) > 0$ implies $\mu(\gamma_Y(B)) > 0$ (because then $Y := \bigcap_{\gamma \in \Gamma} Y_\gamma$ is as desired).

To this end, fix $\gamma \in \Gamma$. We recursively construct a decreasing sequence $(X_n)$ of conull sets and a pairwise disjoint sequence $(B_n)$ of Borel subsets $B_n \subseteq X_n$ as follows. Let $X_0 := X$ and suppose $X_n$ has been constructed. If there is a Borel set $B \subseteq \text{dom}(\gamma_{X_n})$ with $\mu(B) > 0$ and $\mu(\gamma_{X_n}(B)) = 0$, let $B_n$ be one such set with
\[
\mu(B_n) > \frac{1}{2} \sup \{\mu(B) : B \subseteq \text{dom}(\gamma_{X_n}) \text{ Borel with } \mu(\gamma_{X_n}(B)) = 0\}. \tag{2.3}
\]
Then put \( X_{n+1} := X_n \setminus \gamma_{X_n}(B_n) \). Having constructed these sequences of sets, we check that the conull set \( X_\infty := \bigcap_n X_n \) is as desired. Indeed, let \( B \subseteq \text{dom}(\gamma_{X_\infty}) \) be a Borel set with \( \mu(\gamma_{X_\infty}(B)) = 0 \). Then for all \( n \), (2.3) implies \( \mu(B_n) > \frac{1}{2} \mu(B) \), so \( \mu(B) = 0 \) because \( \mu(B_n) \to 0 \), since the \( B_n \) are pairwise disjoint and \( \mu(X) < \infty \). \( \square \)

**Proposition 2.4.** For any countable-to-one null-preserving Borel transformation \( T \) on a standard probability space \((X, \mu)\), there is a conull set \( X' \subseteq X \) such that \( T(X') = X' \) and \( E|_{X'} \) is null-preserving.

**Proof.** Let \( X_0 \subseteq X \) be a conull set given by Lemma 2.2, i.e. \( E_T|_{X_0} \) is null-preserving. Furthermore, because \( T \) is null-preserving, the set \( X_1 := \bigcap_{n \in \mathbb{N}} T^{-n}(X_0) \) is still conull, but we now have \( T(X_1) \subseteq X_1 \). Lastly, again because \( T \) is null-preserving, the \( T \)-image of a conull set is conull, so \( Z := X_1 \setminus T(X_1) \) is null. Hence, \( [Z]_{E_T} \) is also null (because \( E_T|_{X_1} \) is null-preserving), and therefore \( X' := X_1 \setminus [Z]_{E_T} \) is still conull, but now we finally have \( T(X') = X' \). \( \square \)

**Assumption 2.5.** Since all statements in the current paper are modulo null sets, without loss of generality (by Proposition 2.4), we assume that all countable-to-one null-preserving Borel transformations \( T : X \to X \) are surjective and the induced equivalence relations \( E_T \) are null-preserving. Thus, we let \( \rho \) be the corresponding Radon–Nikodym cocycle.

The rest of this section is only used in Section 6. Let \( E \) be a null-preserving countable Borel equivalence relation on a standard probability space \((X, \mu)\). Let \( \operatorname{Erg}_E(X) \) denote the standard Borel space of all \( E \)-ergodic Borel probability measures on \( X \).

**Definition 2.6.** A Borel map \( \epsilon : X \to \operatorname{Erg}_E(X) \) is called an \( E \)-ergodic decomposition of \( \mu \) if for each \( x \in X \), \( \epsilon_x := \epsilon(x) \) is supported on \( \epsilon^{-1}(\epsilon_x) \), and for all Borel sets \( A \subseteq X \),

\[
\mu(A) = \int_X \epsilon_x(A) \, d\mu(x).
\]

It follows from this definition that an \( E \)-ergodic decomposition, if exists, is unique modulo a \( \mu \)-null set. As for the existence, for pmp countable Borel equivalence relations this is due to Farrel and Varadarajan [Far62, Var63] (see also [KM04, Theorem 3.3]), and more generally, it is a theorem of Ditzen for null-preserving equivalence relations [Dit92] (see also [Mil08, Theorem 5.2]). We will only use the existence of an ergodic decomposition for pmp equivalence relations, as well as the following connection with conditional expectation.

**Proposition 2.7** (Conditional expectation via ergodic decomposition). Let \( \epsilon : X \to \operatorname{Erg}_E(X) \) be an \( E \)-ergodic decomposition of \( \mu \). For any \( f \in L^1(X, \mu) \) and \( E \)-invariant Borel set \( B \subseteq X \),

\[
\int_B f \, d\mu = \int_B \int_X f(z) \, d\epsilon_x(z) \, d\mu(x). \tag{2.8}
\]

In particular, for \( \mu \)-a.e. \( x \in X \),

\[
\bar{f}(x) = \int_X f \, d\epsilon_x,
\]

where \( \bar{f} \) is the \( \mu \)-conditional expectation with respect to the \( \sigma \)-algebra of \( E \)-invariant Borel sets.

**Proof.** The “in particular” part follows from the definition and uniqueness of the conditional expectation and the fact that the map \( x \mapsto \int_X f(z) \, d\epsilon_x(z) \) is \( E \)-invariant.
As for the main part, a standard approximation argument gives that for each \( h \in L^1(X, \mu) \),
\[
\int_X h \, d\mu = \int_X \int_X h(z) \, d\epsilon_x(z) \, d\mu(x).
\] (2.9)

**Claim.** For every \( E \)-invariant Borel set \( A \subseteq X \) and \( \mu \)-a.e. \( x \in X \),
\[
\epsilon_x(A) = \begin{cases} 
1 & \text{if } x \in A \\
0 & \text{if } x \notin A. 
\end{cases}
\]

**Proof of Claim.** Indeed, letting \( A' := X_A \setminus A \), where \( X_A := \{ x \in X : \epsilon_x(A) = 1 \} \), we see that
\[
\mu(A') = \int_X \epsilon_x(A') \, d\mu(x) = \int_{X_A} \epsilon_x(A') \, d\mu(x) + \int_{X \setminus X_A} \epsilon_x(A') \, d\mu(x) = 0 + 0 = 0
\]
because for each \( x \in X_A \), \( \epsilon_x(A) = 1 \) and \( A' \cap A = \emptyset \), so \( \epsilon_x(A') = 0 \), and for each \( x \notin X_A \), \( X_A \cap \epsilon^{-1}(\epsilon_x) = \emptyset \), so \( \epsilon_x(X_A) = 0 \), in particular, \( \epsilon_x(A') = 0 \). The same argument, but with \( X \setminus A \) in place of \( A \), shows that \( \mu(A \setminus X_A) = 0 \).

This and (2.9) applied to \( h := f \cdot 1_A \), where \( A \) ranges over \( E \)-invariant Borel subsets of \( X \), imply (2.8). \( \square \)

## 3. Averaging

Let \( X \) be a standard Borel space equipped with a Borel probability measure \( \mu \). Let \( T : X \to X \) be a countable-to-one Borel transformation. Call a set \( \Gamma \) of Borel partial functions \( \gamma : X \to X \) a **complete set of Borel partial right-inverses** of \( T \) if the graphs of the \( \gamma \in \Gamma \) are pairwise disjoint and for each \( x \in X \),
\[
T^{-1}(x) = \{ \gamma(x) : x \in \text{dom}(\gamma) \text{ and } \gamma \in \Gamma \}.
\]

Because \( T \) is countable-to-one, the Luzin–Novikov Uniformization theorem [Kec95, Theorem 18.10] ensures that \( T \) admits a countable complete set \( \Gamma \) of Borel partial right-inverses, and we fix any such a (countable) \( \Gamma \) for the remaining of this section.

For each \( n \in \mathbb{N} \) and \( t \in \Gamma^n \), define the partial function \( t : X \to X \) by recursion on the length of \( t \) as follows: if \( t = \emptyset \), put \( t(x) := x \) for all \( x \in X \). If \( t = \gamma \cdot t', \ x \in \text{dom}(t') \), and \( t'(x) \in \text{dom}(\gamma) \), define \( t(x) := \gamma(t'(x)) \). Otherwise, leave \( t(x) \) undefined.

**Observation 3.1.** If \( S \subseteq \Gamma^n \) for some \( n \in \mathbb{N} \), then the sets \( t(X), \ t \in S \), are pairwise disjoint.

For \( x \in X \) and \( S \subseteq \Gamma^{\leq n} \), define
\[
S \cdot x := \{ t(x) : t \in S \text{ and } x \in \text{dom}(t) \}.
\]

Also, if \( n \in \mathbb{N} \), define
\[
\triangleright_T^n \cdot x := \bigcup_{i \leq n} T^{-i}(x) \ (= \Gamma^{\leq n} \cdot x).
\]

Notice that \( \triangleright_T^n \cdot x \) does not depend on the choice of \( \Gamma \).

Equipping \( X \) with a Borel probability measure \( \mu \), suppose that \( T \) is null-preserving and let \( \rho : E_T \to \mathbb{R}^+ \) denote the corresponding (to \( \mu \) Radon–Nikodym cocycle (see Assumption 2.5).

We will need the following averaging lemmas to prove both the tiling property (Lemma 4.1) and the main backward ergodic theorem (Theorem 4.7).
Lemma 3.2. Let $T : X \to X$ be a countable-to-one null-preserving Borel transformation. For each $n \in \mathbb{N}$, $C \subseteq \Gamma^n$, and $f \in L^1(X, \mu)$,

$$\int_X \sum_{y \in C \cdot x} f(y)\rho(y, x) \, d\mu(x) = \int_{\text{im}(C)} f \, d\mu,$$

where $\text{im}(C) = \bigcup_{t \in C} t(X)$ ($\bigcup$ denotes a disjoint union). In particular,

$$\int_X \sum_{y \in T^{-n}(x)} f(y)\rho(y, x) \, d\mu(x) = \int_X f \, d\mu.$$

Proof. The “in particular” part is just the special case of the main part for $S := \Gamma^n$. As for the main part, it is enough to prove it for non-negative functions, since then the statement for an arbitrary $f \in L^1(X, \mu)$ follows by the decomposition $f = f^+ - f^-$ into positive and negative parts.

$$\int_X \sum_{y \in C \cdot x} f(y)\rho(y, x) \, d\mu(x) = \int_X \sum_{t \in C} \mathbb{1}_{\text{dom}(t)}(x)f(t(x))\rho(t(x), x) \, d\mu(x)$$

\[\text{because } f \geq 0\]

$$= \sum_{t \in C} \int_X \mathbb{1}_{\text{dom}(t)}(x)f(t(x))\rho(t(x), x) \, d\mu(x)$$

\[\text{by Eq. (2.1)}\]

$$= \sum_{t \in C} \int_{\text{im}(t)} f(x) \, d\mu(x)$$

$$= \int_{\text{im}(C)} f(x) \, d\mu(x). \quad \square$$

Corollary 3.3. Let $T : X \to X$ be a countable-to-one null-preserving Borel. For any $f \in L^1(X, \mu)$ and for a.e. $x \in X$, for all $N \in \mathbb{N}$ and all $S \subseteq \Gamma^N$,

$$\sum_{y \in S \cdot x} |f(y)|\rho(y, x) \, d\mu(x) < \infty.$$

Proof. By Lemma 3.2,

$$\int_X \sum_{y \in T^{-n}(x)} |f(y)|\rho(y, x) \, d\mu(x) \, d\mu = \|f\|_1 < \infty,$$

which implies $\sum_{y \in T^{-n}(x)} |f(y)|\rho(y, x) < \infty$ a.e. $x \in X$. Switching the quantifiers, we get that for a.e. $x \in X$, for all $n \in \mathbb{N}$, $\sum_{y \in T^{-n}(x)} |f(y)|\rho(y, x) \, d\mu(x) < \infty$. Thus,

$$\sum_{y \in S \cdot x} |f(y)|\rho(y, x) \, d\mu(x) \leq \sum_{y \in \mathbb{N}^N \cdot x} |f(y)|\rho(y, x) \, d\mu(x) = \sum_{n \leq N} \sum_{y \in T^{-n}(x)} |f(y)|\rho(y, x) < \infty. \quad \square$$

Lemma 3.4. Let $T : X \to X$ be a countable-to-one null-preserving Borel. For any $f \in L^1(X, \mu)$ and $N \in \mathbb{N}$,

$$\int_X f(x) \, d\mu(x) = \int_X \frac{1}{N+1} \sum_{y \in \mathbb{N}^N \cdot x} f(y)\rho(y, x) \, d\mu(x).$$
Proof. Applying Lemma 3.2 for each $n \in \mathbb{N}$, we get:

\[
\int_X f(x) \, d\mu(x) = \int_X \frac{N + 1}{N + 1} f(x) \, d\mu(x)
= \frac{1}{N + 1} \int_X \sum_{n=0}^{N} \sum_{y \in T^{-n}(x)} f(y) \rho(y, x) \, d\mu(x)
= \frac{1}{N + 1} \sum_{y \in T^{N}x} f(y) \rho(y, x) \, d\mu(x).
\]

□

Using this, we get a bound on the $L^p$-norms of averages over subsets of $\Gamma^{\leq N}$.

**Corollary 3.5.** Let $T : X \to X$ be a countable-to-one null-preserving Borel transformation and $\emptyset \in S \subseteq \Gamma^{\leq N}$ for some $N \in \mathbb{N}$. For any $f \in L^p(X, \mu)$ and $p \geq 1$,

\[
\| x \mapsto \left( \frac{|S \cdot x|^p}{N + 1} \right)^{\frac{1}{p}} A^p_f[S \cdot x] \|_p \leq \| f \|_p.
\]

Proof. By Jensen’s inequality

\[
\int_X \frac{|S \cdot x|^p}{N + 1} (A^p_f[S \cdot x])^p \, d\mu(x) \leq \int_X \frac{|S \cdot x|^p}{N + 1} A^p_f[S \cdot x] \, d\mu(x)
= \int_X \frac{1}{N + 1} \sum_{y \in S \cdot x} |f(y)|^p \rho(y, x) \, d\mu(x)
\leq \int_X \frac{1}{N + 1} \sum_{y \in T^N \cdot x} |f(y)|^p \rho(y, x) \, d\mu(x)
\]

[by Lemma 3.4] $\| f \|_p.$ □

**Lemma 3.6.** For a countable-to-one $\mu$-preserving Borel transformation $T : X \to X$, for a.e. $x \in X$, and for each $n \in \mathbb{N}$, $|T^{-n}(x)|^p = 1$. In particular, for each $N \in \mathbb{N}$, $|T^N \cdot x|^p = N + 1$ a.e.

Proof. The second statement is immediate from the first. For the first statement, we may switch the quantifiers, i.e. prove that for each $n \in \mathbb{N}$, the formula holds a.e. Induction on $n$ using the cocycle identity or simply the fact that $T^n$ is also $\mu$-preserving shows that it is enough to prove the statement for $n = 1$.

To this end, we show that $A := \{ x \in X : |T^{-1}(x)|^p > 1 + \varepsilon \}$ is null for each $\varepsilon > 0$. This implies that $\{ x \in X : |T^{-1}(x)|^p > 1 \}$ is null and an analogous argument shows that
\{x \in X : |T^{-1}(x)|^x_\rho < 1\} is null as well. Since \(T\) is \(\mu\)-preserving,

\[
\mu(A) = \mu(T^{-1}(A)) = \mu \left( \bigcup_{\gamma \in \Gamma} \gamma(A \cap \text{dom}(\gamma)) \right)
\]

\[
= \sum_{\gamma \in \Gamma} \mu(\gamma(A \cap \text{dom}(\gamma)))
\]

[by Eq. (2.1)]

\[
= \sum_{\gamma \in \Gamma} \int_{A \cap \text{dom}(\gamma)} \rho(\gamma(x), x) \, d\mu(x)
\]

\[
= \int_A \sum_{\gamma \in \Gamma} 1_{\text{dom}(\gamma)}(x) \rho(\gamma(x), x) \, d\mu(x)
\]

\[
= \int_A \sum_{y \in T^{-1}(x)} \rho(y, x) \, d\mu(x)
\]

\[
= \int_A |T^{-1}(x)|^x_\rho \, d\mu(x) \geq \mu(A)(1 + \varepsilon).
\]

This implies that \(A\) is null, as desired. \(\square\)

Applying Lemma 3.6 to Lemma 3.4 immediately yield the following.

**Corollary 3.7** (Averaging over complete trees). Let \(T : X \to X\) be a countable-to-one \(\mu\)-preserving Borel transformation. For any \(f \in L^1(X, \mu)\) and \(N \in \mathbb{N}\),

\[
\|A_\rho f \|_{\text{\(\geq\)} N} \cdot x \|_p \leq \int_X A_\rho^p \|[x, 0^-] \|_p \, d\mu(x).
\]

**Remark 3.8.** Corollary 3.7 fails when we replace \(\geq N\) \cdot x\) with arbitrary subsets of the back-orbit of \(x\), even trees (as in Section 4.1). We may observe this by looking at indicator functions of the ranges of the \(\gamma_n\). For example, if \(T\) is the shift map on \((2^N, \{\frac{1}{2}, \frac{1}{2}\}^{\mathbb{N}}),\) and \(f := 1_{\{x_0 = 0\}},\) then \(\int f \, d\mu = \frac{1}{2},\) but \(\int A_\rho f \left[\{x, 0^-\} \right] \, d\mu = \frac{2}{3}.\)

Another application of Lemma 3.6, this time to Corollary 3.5, yields the following bound.

**Corollary 3.9.** Let \(T : X \to X\) be a countable-to-one \(\mu\)-preserving Borel transformation. For any \(f \in L^p(X, \mu), p \geq 1,\) and any \(N \in \mathbb{N}\),

\[
\|A_\rho f \|_{\text{\(\geq\)} N} \cdot x \|_p \leq \|f\|_p.
\]

### 4. The tiling property and the backwards ergodic theorem

Throughout, let \((X, \mu)\) be a standard probability space and let \(T : X \to X\) be a countable-to-one \(\mu\)-preserving Borel transformation, so by Assumption 2.5, \(T\) is surjective and \(E_T\) is null-preserving. Let \(\rho : E_T \to \mathbb{R}^+\) be the Radon–Nikodym cocycle corresponding to \(\mu\). Finally, let \(\Gamma\) be a complete set of Borel partial right-inverses of \(T\).

#### 4.1. The tiling property.

We now prove the tiling property and deduce our more general backward pointwise ergodic theorem (Theorem 4.7) from it.

For a set \(I\) (which will typically be a countable complete set of Borel right-inverses of a transformation \(T\)), let \(\mathcal{T}_I \subseteq \mathcal{P}(I^{<\mathbb{N}})\) be the set of nonempty set-theoretic (but right-rooted)
trees on $I$ of finite height, where $I^{<\mathbb{N}}$ is the set of all finite sequences of elements of $I$. More precisely, for each $S \subseteq I^{<\mathbb{N}}$,

$$S \in \mathcal{T}_I :<\leftrightarrow S \subseteq \mathcal{P}(I^{\leq\mathbb{N}}) \text{ for some } N, \emptyset \in S,$$

and for each $t \in S$, all of its terminal subsequences are also in $S$.

For each $S \in \mathcal{T}_I$, denote by $h(S)$ the height of the tree $S$, i.e. the least $N \in \mathbb{N}$ such that $S \subseteq I^{\leq\mathbb{N}}$.

**Lemma 4.1** (Tiling property). Let $(X, \mu)$ be a standard probability space, $T : X \to X$ be a countable-to-one $\mu$-preserving Borel transformation, so by Assumption 2.5, $T$ is surjective and $E_T$ is null-preserving. Let $\Gamma$ be a complete set of Borel partial right-inverses of $T$, and let $\rho : E_T \to \mathbb{R}^+$ be the Radon–Nikodym cocycle corresponding to $\mu$.

Then for any measurable function $x \mapsto S_x : X \to \mathcal{T}_1$ and $\varepsilon > 0$ there is $N \in \mathbb{N}$ such that for at least $(1 - \varepsilon)$-measured set of $x \in X$, $\triangleright_N \cdot x$ can be covered, up to an $\varepsilon \cdot \rho$-fraction, by disjoint tiles of the form $S_y \cdot y$ for $y \in \triangleright_N \cdot x$.

**Proof.** Let $L$ be large enough so that the set

$$B := \{x \in X : h(S_x) \geq L\}$$

has measure less than $\frac{\varepsilon^2}{2}$. Fix $N$ large enough so that $\frac{L}{N} < \frac{\varepsilon}{2}$. By Lemma 3.6, we may assume that for each $x \in X$ and $n \in \mathbb{N}$, $\sum_{y \in T^{-n}(x)} (\rho(y, x)) = 1$, so $|\triangleright_N^{-L}(x)|_\rho > \frac{\varepsilon}{2} |\triangleright_N \cdot x|_\rho$. Thus, there is no harm in leaving $\triangleright_N \cdot x \setminus \triangleright_N^{-L}(x)$ untiled.

We claim that for all but less than $\varepsilon$-measured set of $x \in X$, less than $\frac{\varepsilon}{2} \rho$-fraction of $y \in \triangleright_N \cdot x$ are in $B$, i.e. the set

$$C := \left\{x \in X : A_{\triangleright_N}^\rho [\triangleright_N \cdot x] \geq \frac{\varepsilon}{2}\right\},$$

has measure less than $\varepsilon$. Indeed:

$$\frac{\varepsilon^2}{2} > \mu(B) = \int_X 1_B(x) \, d\mu(x) \quad \left[\text{by Corollary 3.7} \right] = \int_X A_{\triangleright_N}^\rho [\triangleright_N \cdot x] \, d\mu(x) \geq \int_C A_{\triangleright_N}^\rho [\triangleright_N \cdot x] \, d\mu(x) \geq \frac{\varepsilon}{2} \mu(C).$$

So we just need to fix $x \in X \setminus C$ and tile the set $\triangleright_N \cdot x$ up to an $\varepsilon \cdot \rho$-fraction. We do this by the following straightforward algorithm: if there is $n \leq N$ with a $y \in T^{-n}(x)$ that is not covered by a tile yet and $\delta_{S_y} \cdot y \subseteq \triangleright_N(x)$, take the least such $n$ and for each such $y \in T^{-n}(x)$, place the tiles $S_y \cdot y$; repeat this until there is no such $n$. Once this process terminates, the only points that are not covered by a tile must belong to either $B$ or $\triangleright_N \cdot x \setminus \triangleright_N^{-L}(x)$, so they comprise at most $\varepsilon \cdot \rho$-fraction of $\triangleright_N \cdot x$. □

4.2. **T-recurrence.** Here, we recall some ergodic-theoretic terminology and basic facts, which are used in Section 4.3. A set $W \subseteq X$ is called $T$-wandering if the sets $T^{-n}(W)$, $n \in \mathbb{N}$, are pairwise disjoint. Because the measures of the sets $T^{-n}(W)$ are all equal and $\mu$ is a probability measure, we have
Observation 4.2. \( T \) is conservative, i.e., every \( T \)-wandering measurable set is null.

For a set \( U \subseteq X \), let
\[
[U]^+_T := \bigcup_{n \in \mathbb{N}^+} T^n(U), \quad [U]^{-1}_T := \bigcup_{n \in \mathbb{N}^+} T^{-n}(U), \quad [U]_{T^{-1}} := \bigcup_{n \in \mathbb{N}} T^{-n}(U).
\]
Abusing notation, we write \([x]_T^+\) when \( U = \{x\} \). Note that for any set \( U \), the set \( V := X \setminus [U]_{T^{-1}} \) is closed under \( T \), i.e., \( T(V) \subseteq V \).

Call a set \( U \subseteq X \) \( T \)-recurrent if for every \( x \in U \), \([x]_T^+ \cap U \neq \emptyset\); equivalently, \( U \setminus [U]_{T^{-1}}^+ = \emptyset \).

Lemma 4.3. Every Borel set \( U \subseteq X \) is \( T \)-recurrent a.e. In fact, there is a subset \( U' \subseteq U \) that is conull in \( U \) such that for every \( x \in [U']_{E_T} \), \([x]_T^+ \cap U' \neq \emptyset \).

Proof. The set \( U'' := \left\{ x \in U : [x]_T^+ \cap U = \emptyset \right\} \) is \( T \)-wandering and hence null. Then \([U'']_{T^{-1}} \) is also null, and it is easy to check that \( U' := U \setminus [U'']_{T^{-1}} \) is as desired.

In light of Lemma 4.3, we may assume that all positively measured sets that come up are \( T \)-recurrent.

Lemma 4.4. Every Borel set \( U \subseteq X \) with the property that \( T(U) \subseteq U \) is such that \([U]_{E_T} = U \) off of an invariant null set.

Proof. Put \( V := [U]_{E_T} \setminus U \). Then since \( T(U) \subseteq U \), \( V \) is nowhere recurrent, i.e., for all \( x \in V \), there are only finitely many \( n \) with \( T^n(x) \in V \). Hence, by Lemma 4.3, \( V \) is null, and since \( E_T \) is null-preserving, so is \([V]_{E_T} \). Therefore, \( U \) is \( E_T \)-invariant off of the invariant null set \([V]_{E_T} \).

We say that the periodic part of \( T \) is the subset \( \{ x \in X : \exists n < m \in \mathbb{N} : T^n(x) = T^m(x) \} \).

Lemma 4.5. Off of a null set, \( T \) is bijective on its periodic part.

Proof. Let \( V \) be the periodic part of \( T \), and let \( U := \{ x \in X : \exists n \in \mathbb{N} \setminus \{0\} : T^n(x) = x \} \). Then \([U]_{E_T} = V \). Notice that \( V \setminus U \) is nowhere \( T \)-recurrent, hence null, and that \( T|_U \) is bijective.

4.3. Pointwise ergodic theorem along trees. For \( x \in X \), let \( \mathcal{T}_x \) denote the collection of subtrees of the graph \( \text{graph}(T) \) of \( T \) of finite height rooted at \( x \) and directed towards \( x \). More precisely, \( S_x \in \mathcal{T}_x \) if and only if \( x \in S_x \), \( S_x \subseteq \bigcup_{i=0}^n T^{-i}(x) \) for some \( n \in \mathbb{N} \), and if \( y \in S_x \) and \( y \neq x \) then \( T(y) \in S_x \). Notice that if \( \Gamma \) is a complete set of Borel partial right-inverses of \( T \), \( S_x \in \mathcal{T}_x \) exactly when \( S_x = S \cdot x \) for some \( S \in \mathcal{T}_T \). With this in mind, we use graph-theoretic trees (\( S_x \subseteq \text{graph}(T) \)) and set-theoretic trees (\( S \subseteq \mathcal{T}^{<\mathbb{N}} \)) interchangeably in the rest of the paper.

For \( f \in L^1(X, \mu) \) and \( x \in X \), we write
\[
\lim_{S_x \in \mathcal{T}_x} A_f^x[S_x] = L
\]
to mean that for any \( \varepsilon > 0 \), there is \( w > 0 \) such that for any \( S_x \in \mathcal{T}_x \), \( |S_x|^\rho > w \) implies \( |A_f^x[S_x] - L| < \varepsilon \). Similarly, we write \( \lim sup_{S_x \in \mathcal{T}_x} A_f^x[S_x] = L \) if for any \( \varepsilon > 0 \) there are \( S_x \in \mathcal{T}_x \) with arbitrarily large \( |S_x|^\rho \) for which \( A_f^x[S_x] > L - \varepsilon \), and \( \lim inf_{S_x \in \mathcal{T}_x} A_f^x[S_x] = L \) if for any \( \varepsilon > 0 \) there are \( S_x \in \mathcal{T}_x \) with arbitrarily large \( |S_x|^\rho \) for which \( A_f^x[S_x] < L + \varepsilon \).
In the examples in Section 5 and Section 6, it will be helpful to think of these limits in terms of a certain natural complete set $\Gamma$ of Borel partial right-inverses of $T$. In this case, for $f \in L^1(X, \mu)$ and $x \in X$, we write
\[
\lim_{S \in \mathcal{T}_x} A^\rho_{f}[S \cdot x] = L
\]
to mean that for any $\varepsilon > 0$, there is $w > 0$ such that for any $S \in \mathcal{T}_x$, $|S \cdot x|_\rho > w$ implies $|A^\rho_{f}[S \cdot x] - L| < \varepsilon$. We define $\limsup_{S \in \mathcal{T}_x} A^\rho_{f}[S \cdot x] = L$ and $\liminf_{S \in \mathcal{T}_x} A^\rho_{f}[S \cdot x] = L$, analogously.

We will ultimately show that $\lim_{S_x \in \mathcal{T}_x} A^\rho_{f}[S_x]$ exists for $\mu$-a.e. $x \in X$; in particular, the pointwise limit of averages over $v^\rho_T \cdot x$ exists a.e. In order to prove the former, we first need that $\limsup_{S_x \in \mathcal{T}_x} A^\rho_{f}[S_x]$ and $\liminf_{S_x \in \mathcal{T}_x} A^\rho_{f}[S_x]$ are $E_T$-invariant modulo an invariant null set.

**Lemma 4.6.** Let $(X, \mu)$ be a standard probability space, $T : X \to X$ be an aperiodic countable-to-one $\mu$-preserving Borel transformation, so by Assumption 2.5, $T$ is surjective and $E_T$ is null-preserving. Let $\rho : E_T \to \mathbb{R}^+$ be the Radon–Nikodym cocycle corresponding to $\mu$. Then for any $f \in L^1(X, \mu)$, for a.e. $x \in X$, for all $S_x \in \mathcal{T}_x$, $A^\rho_{f}[S_x] < \infty$, and the functions
\[
f^*(x) := \limsup_{S_x \in \mathcal{T}_x} A^\rho_{f}[S_x] \quad \text{and} \quad f_*(x) := \liminf_{S_x \in \mathcal{T}_x} A^\rho_{f}[S_x]
\]
are $T$-invariant a.e. (i.e. off of an invariant null set).

**Proof.** That for a.e. $x \in X$ and each $S_x \in \mathcal{T}_x$, $A^\rho_{f}[S_x] < \infty$ is the content of Corollary 3.3. As for invariance, it is enough to show that $f^*$ is $T$-invariant as $f_* = -(f)^*$. For that, it is enough to show that for each $a \in \mathbb{Q}$, the set
\[
X_{\geq a} := \{ x \in X : f^*(x) \geq a \}
\]
is $E_T$-invariant, modulo a null set. Fix $a \in \mathbb{Q}$. By Lemma 4.4, we just need to show $T(X_{\geq a}) \subseteq X_{\geq a}$. For this, it suffices to show that for all $x \in X$, $f^*(x) \leq f^*(T(x))$.

Fix $x \in X$. Intuitively, since $T(x)$ is only one point, we can add it to large sets $S \in \mathcal{T}_x$ without having much impact on the weighted average. Hence, $f^*(x) \leq f^*(T(x))$. To see this more formally, first assume $f^*(x) \in \mathbb{R}$. Fix $\varepsilon > 0$ and a weight $w > 0$. We need to find $S_T(x) \in \mathcal{T}_T(x)$ such that $|S_T(x)|_\rho > \rho(T(x), x)w$ (equivalently, $|S_T(x)|^{T(x)}_\rho > w$) and $A^\rho_{f}[S_T(x)] \geq f^*(x) - \varepsilon$. To this end, let $S'_x \in \mathcal{T}_x$ such that $|S'_x|_\rho > \max \left\{ \rho(T(x), x)w, \frac{4|f^*(x)||T(x)|_\rho^2}{\varepsilon}, \frac{4|f^*(x)||T(x)|_\rho^2}{\varepsilon} \right\}$ and $A^\rho_{f}[S'_x] \geq f^*(x) - \frac{\varepsilon}{2}$. Put $S_T(x) := \{ T(x) \} \cup S'_x$ (hence, $S_T(x) \in \mathcal{T}_T(x)$). Thus, we have
\[
A^\rho_{f}[S_T(x)] \geq \frac{|S'_x|_\rho |A^\rho_{f}[S'_x]|}{|S_T(x)|_\rho^2} A^\rho_{f}[S'_x] + \frac{|T(x)|_\rho^2}{|S_T(x)|_\rho^2} f(T(x))
\]
\[
\geq \left( 1 - \frac{\varepsilon}{4|f^*(x)|} \right) \left( f^*(x) - \frac{\varepsilon}{2} \right) - \frac{\varepsilon}{4|f^*(x)|} |f(T(x))|
\]
\[
\geq f^*(x) - \frac{3\varepsilon}{4} - \frac{\varepsilon}{4} = f^*(x) - \varepsilon.
\]
The case where $f^*(x) = \infty$ is handled similarly. Thus, $f^*(x) \leq f^*(T(x))$. \qed

**Theorem 4.7** (Backward pointwise ergodic theorem along trees). Let $(X, \mu)$ be a standard probability space, $T : X \to X$ be an aperiodic countable-to-one $\mu$-preserving Borel transformation, so by Assumption 2.5, $T$ is surjective and $E_T$ is null-preserving. Let $\rho : E_T \to \mathbb{R}^+$ be
the Radon–Nikodym cocycle corresponding to \( \mu \). Then for any \( f \in L^1(X, \mu) \), for a.e. \( x \in X \), for all \( S_x \in \mathcal{T}_x \), \( A^0_{f^*}[S_x] < \infty \), and

\[
\lim_{S_x \in \mathcal{T}_x} A^0_f[S_x] = \overline{f}(x),
\]

where \( \overline{f} \) is the conditional expectation of \( f \) with respect to the \( \sigma \)-algebra of \( T \)-invariant Borel sets.

**Proof.** By Lemma 4.6, for a.e. \( x \in X \) and each \( S_x \in \mathcal{T}_x \), \( A^0_{f^*}[S_x] < \infty \), and \( f^* \) and \( f_* \) are \( T \)-invariant. By replacing \( f \) with \( f - \overline{f} \), we may assume without loss of generality that \( \overline{f} = 0 \). We will show that \( f^* \leq 0 \) a.e., and an analogous argument shows \( f_* \geq 0 \) a.e.

Assume by way of contradiction that \( f^* > 0 \) on a positively measured (necessarily \( T \)-invariant) set, restricting to which we might as well assume that \( f^*(x) > 0 \) for all \( x \in X \). Put \( g := \min \{ \frac{c}{2}, 1 \} \), so \( 0 < g \leq 1 \), and \( g \in L^1(X, \mu) \). Put \( c := \int g \, d\mu > 0 \). Fix a complete set \( \Gamma \) of Borel partial right-inverses of \( T \), and let \( \mathcal{T}_1^f \subseteq \mathcal{T}_f \) denote the collection of finite trees in \( \mathcal{T}_f \). Notice that for each \( x \in X \), since \( \sum_{y \in S_x} f(y) \rho(y, x) \) converges absolutely for each \( S_x \in \mathcal{T}_x \),

\[
f^*(x) = \limsup_{S \in \mathcal{T}_1^f} A^0_f[S \cdot x] = \limsup_{S \in \mathcal{T}_1^f} A^0_f[S \cdot x].
\]

Fix an enumeration \( \{ S_n \} \) of \( \mathcal{T}_1^f \), and define \( \ell : X \to \mathcal{T}_1^f \) by \( x \mapsto \) the least (according to the enumeration) \( S \in \mathcal{T}_1^f \) such that \( A^0_f[S \cdot x] > g(x) \) (equivalently, \( A^0_{f^*}[S \cdot x] > 0 \)).

Fix \( \delta > 0 \) small enough so that for any measurable \( Y \subseteq X \), \( \mu(Y) < \delta \) implies \( \int_Y (f - g) \, d\mu > -\varepsilon \frac{c}{3} \), and let \( M \in \mathbb{N} \) be large enough so that the set \( Y := f^{-1}(-M, \infty) \) has measure at least \( 1 - \varepsilon \).

The tiling property (Lemma 4.1) applied to the function \( \ell \) with \( \varepsilon := \frac{1}{2(M+1)\frac{c}{3}} \) gives \( N \in \mathbb{N} \) such that \( \mu(Z) \geq 1 - \varepsilon \), where \( Z \) is the set of all \( x \in X \) such that at least \( 1 - \varepsilon \) \( \rho \)-fraction of \( \triangledown^n_T \cdot x \) is partitioned into sets of the form \( \ell(y) \cdot y \).

**Claim.** For each \( x \in Z \), \( A^0_{\mathbb{1}_Y(f-g)}[\triangledown^n_T \cdot x] \geq -(M + 1)\varepsilon \).

**Proof of Claim.** By the definition of \( Z \), on a subset \( B \subseteq \triangledown^n_T \cdot x \) that occupies at least \( 1 - \varepsilon \) \( \rho \)-fraction of \( \triangledown^n_T \cdot x \), the \( \rho \)-average of \( f - g \) is positive, and hence that of \( \mathbb{1}_Y(f - g) |_B \) is non-negative. On the remaining set \( \triangledown^n_T \cdot x \setminus B \), the function \( \mathbb{1}_Y(f - g) \) is at least \(-M + 1\), by the definition of \( Y \), and hence so is its \( \rho \)-weighted average. Thus, the \( \rho \)-weighted average of \( \mathbb{1}_Y(f - g) \) on the entire \( \triangledown^n_T \cdot x \) is at least \(-(M+1)\varepsilon \).

Now we compute using this claim and Corollary 3.7:

\[
\int_Y (f - g) \, d\mu = \int_X A^0_{\mathbb{1}_Y(f-g)}[\triangledown^n_T \cdot x] \, d\mu(x) \\
= \int_Z A^0_{\mathbb{1}_Y(f-g)}[\triangledown^n_T \cdot x] \, d\mu(x) + \int_{X \setminus Z} A^0_{\mathbb{1}_Y(f-g)}[\triangledown^n_T \cdot x] \, d\mu(x) \\
\geq -(M + 1)\varepsilon - (M + 1)\varepsilon = -2(M + 1)\varepsilon = -\frac{c}{3}.
\]
This gives a contradiction:

\[ 0 = \int_X TF \, d\mu = \int_X f \, d\mu = c + \int_X (f - g) \, d\mu = c + \int_Y (f - g) \, d\mu + \int_{X \setminus Y} (f - g) \, d\mu > c - \frac{c}{3} - \frac{c}{3} > 0. \]

Remark 4.8. It is worth explicitly pointing out sequences of trees in \( T_F \) whose \( \rho \)-weights go to infinity regardless of the base point \( x \). Such is the sequence \( v_1^n \); indeed, by Lemma 3.6, \(|v_1^n \cdot x|_\rho^n = n + 1 \) for a.e. \( x \in X \). More generally, this is true for sequences \( (S_n) \) of trees that contain triangles rooted at a bounded level and whose heights tend to infinity. By this we mean that there is \( h_0 \in \mathbb{N} \) such that for each \( h \in \mathbb{N} \) there is \( N \in \mathbb{N} \) such that for all \( n \geq N \), \( S_n \) contains a triangle of height at least \( h \) rooted at most at the height \( h_0 \).

Remark 4.9. By Lemma 4.5, the periodic part of \( T \) is bijective mod null, so by Assumption 2.5, we can assume the periodic part is bijective. Consequently, \( E_T \) is measure-preserving on the periodic part, so the backward averages are unweighted (as in the standard pointwise ergodic theorem for \( Z \)), although the sets will not be trees in the graph of \( T \) since the graph of \( T \) on this part consists of finite cycles. We can drop the aperiodicity assumption, but we would need to take the averages over set-theoretic trees (i.e., over \( S \cdot x \) where \( S \in T_F \) for some complete set \( \Gamma \) of Borel partial right-inverses of \( T \)) instead of graph-theoretic subtrees of the graph of \( T \).

If \( T \) is as in Theorem 4.7, then for \( \mu \)-a.e. \( x \in X \), the set \( \left\{ \frac{1}{n} \sum_{i<n} f(T^i(x)) : n \in \mathbb{N} \right\} \) of forward averages is bounded (since this is a convergent sequence). On the other hand, it is not true in general that if a function \( g \) of \( S_x \), \( S_x \in T_x \), converges as \( |S_x|^\|_\rho \to \infty \), then the set \( \{g(S_x) : S_x \in T_x\} \) is bounded. In our case, however, this does hold; i.e., the set of backward averages \( \{A_f[S_x] : S_x \in T_x\} \) is bounded for each \( x \) for which the conclusion of Theorem 4.7 holds. This is demonstrated in the following corollary, which will be used in the proof of Theorem 1.1, but is also interesting in its own right.

Corollary 4.10. Let \((X, \mu)\) be a standard probability space, \( T : X \to X \) be an aperiodic countable-to-one \( \mu \)-preserving Borel transformation, \( \rho : E_T \to \mathbb{R}^+ \) be the Radon–Nikodym cocycle corresponding to \( \mu \), and \( f \in L^1(X, \mu) \). Then for a.e. \( x \in X \), the set of backward averages \( \{A_f[S_x] : S_x \in T_x\} \) is bounded.

Proof. Since \(|A_f[S_x]| \leq A_{|f|}[S_x]| \), we may assume without loss of generality that \( f \geq 0 \). It suffices to prove this for each \( x \in X \) which satisfies the conclusion of Theorem 4.7. Fix such an \( x \), and let \( M \in \mathbb{N} \) be large enough so that \(|S_x|^\|_\rho \geq M \) implies \(|A_f[S_x] - \bar{f}(x)| < 1 \). Therefore, \( \{A_f[S_x] : |S_x|^\|_\rho \geq M \in T_x\} \) is bounded, so we just need to show that \( \{A_f[S_x] : |S_x|^\|_\rho < M \in T_x\} \) is bounded.

Fix \( S_x \in T_x \) with \(|S_x|^\|_\rho < M \). We claim \( A_f[S_x] < (2M + 1)(\bar{f}(x) + 1) \). To see this, let \( y \notin S_x \) such that \( T(y) \in S_x \). Let \( N \in \mathbb{N} \) so that \( M \leq \rho(y, x)N < M + 1 \). Hence, \( M \leq |\Delta_N^{-1}(y)|_\rho \geq M + 1 \) because \(|\Delta_N^{-1}(y)|_\rho \geq |\Delta_N^{-1}(y)|_\rho \geq \rho(y, x)N \).

Then \( S_x' := S_x \cup \Delta_N^{-1}(y) \in T_x \), and \( M < |S_x'| < 2M + 1 \). Hence, we have that \( 0 \leq A_f[S_x] = \frac{|S_x'|_\rho}{|S_x|_\rho} A_f[S_x'] - \frac{|\Delta_N^{-1}(y)|_\rho}{|S_x|_\rho} A_f[\Delta_N^{-1}(y)] \leq \frac{|S_x'|_\rho}{|S_x|_\rho} A_f[S_x'] < (2M + 1)(\bar{f}(x) + 1). \)

\( \square \)
4.4. Ergodic theorems along special sequences of trees. Besides pointwise convergence, we also get convergence in $L^p$ along the sequence of complete trees, i.e., the sets $\triangledown^n_T$. This is the content of Corollary 1.3, which we restate and prove here. We also remark afterwards that the theorem holds for other special sequences of trees as well.

**Corollary 4.11** (Backward pointwise ergodic theorem along complete trees). Let $(X, \mu)$ be a standard probability space, $T : X \to X$ be an aperiodic countable-to-one $\mu$-preserving Borel transformation, so by Assumption 2.5, $T$ is surjective and $E_T$ is null-preserving. Let $\rho : E_T \to \mathbb{R}^+$ be the Radon–Nikodym cocycle corresponding to $\mu$. Then for any $f \in L^1(X, \mu)$, for $\mu$-a.e. $x \in X$, for all $n \in \mathbb{N}$, $A^p_f[\triangledown^n_T \cdot x] < \infty$, and

$$\lim_{n \to \infty} A^p_f[\triangledown^n_T \cdot x] = \overline{f}(x),$$

where $\overline{f}$ is the conditional expectation of $f$ with respect to the $\sigma$-algebra of $T$-invariant Borel sets. Furthermore, if $f \in L^p(X, \mu)$, 1 ≤ $p < \infty$, then $\overline{f} \in L^p$, and $A^p_f[\triangledown^n_T \cdot x] \to_{L^p} \overline{f}(x)$.

**Proof.** The pointwise convergence follows immediately from Theorem 4.7 by considering the sequence $\Gamma^\leq n \subset \mathcal{T}_T$, since $\Gamma^\leq n \cdot x = \triangledown^n_T \cdot x$ and $|\triangledown^n_T \cdot x|_\rho = n + 1 \to \infty$ for each $x \in X$.

Now assume $f \in L^p(X, \mu)$ for 1 ≤ $p < \infty$. Since conditional expectation is an $L^p$ contraction (see [Dur19, Theorem 4.1.11]), $\overline{f} \in L^p(X, \mu)$. Since $L^p(X, \mu) \subseteq L^1(X, \mu)$, $f \in L^1(X, \mu)$, so by the above argument we have that for $\mu$-a.e. $x \in X$, $\lim_{n \to \infty} A^p_f[\triangledown^n_T \cdot x] = \overline{f}(x)$.

If $f \in L^\infty(X, \mu)$, i.e., $|f| \leq M$ for some $M < \infty$, then for each $n$, $|A^p_f[\triangledown^n_T \cdot x]| \leq M$, so by the dominated convergence theorem, $|A^p_f[\triangledown^n_T \cdot x] - \overline{f}(x)|_p \to L^p 0$. For a general $f \in L^p(X, \mu)$, let $f_k$ be a sequence of bounded $L^p$ functions such that $\|f - f_k\|_p \to 0$ (hence $\|\overline{f} - f_k\|_p = \|\overline{f} - f_k\|_p \to 0$ as well since conditional expectation is an $L^p$ contraction).

Fix $\varepsilon > 0$, and let $k$ be large enough so that $\|f - f_k\|_p < \frac{\varepsilon}{3}$ and $\|\overline{f} - f_k\|_p < \frac{\varepsilon}{3}$. Since the theorem holds for $f_k$, let $N$ be large enough so that for $n > N$, $|A^p_f[\triangledown^n_T \cdot x] - f_k|_p < \frac{\varepsilon}{3}$. So for $n > N$,

$$\|A^p_f[\triangledown^n_T \cdot x] - \overline{f}\|_p \leq \|A^p_f[\triangledown^n_T \cdot x] - A^p_{f_k}[\triangledown^n_T \cdot x]\|_p + \|A^p_{f_k}[\triangledown^n_T \cdot x] - \overline{f}\|_p$$

$$\leq \|A^p_{f - f_k}[\triangledown^n_T \cdot x]\|_p + \|A^p_{f_k}[\triangledown^n_T \cdot x] - f_k\|_p + \|f_k - \overline{f}\|_p$$

$$\leq \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} = \varepsilon.$$

Hence, $|A^p_f[\triangledown^n_T \cdot x] - \overline{f}|_p \to 0$, as desired. \hfill \Box

**Remark 4.12.** By Corollary 3.5, Corollary 4.11 still holds when instead of $\triangledown^n_T$ we take any sequence $(S_n) \subseteq \mathcal{T}_T$ such that $h(S_n) \to \infty$ and there is a constant $c > 0$ such that $\frac{|S_n|_\rho}{h(S_n)} \geq c$ for all $x \in X$ (in the case of $\triangledown^n_T$, $c = 1$ works).

**Open Question 4.13.** The authors have not worked out whether the conclusion of Corollary 4.11 holds for general sequences of trees whose weights (as functions of $x$) tend to infinity pointwise.
5. Applications to shift maps

An example of a countable-to-one Borel transformation is the shift function \( s : I^\mathbb{N} \to I^\mathbb{N} \), for some countable set \( I \), where \( s(x) := (x(1 + n))_{n \in \mathbb{N}} \). See Fig. 3 for the depiction of \( s \) for \( I := 2 := \{0, 1\} \).

![Shift on 2^N](image)

**Figure 3.** Shift on \( 2^\mathbb{N} \)

With the exception of the first example, the measures on \( I^\mathbb{N} \) discussed in this section will be Markov chains. Indeed, Markov chains on the state space \( I \) provide a large class of Borel probability measures on \( I^\mathbb{N} \), including shift-invariant ones (which, as verified below, are exactly those with a stationary initial distribution). Thus, Theorem 4.7 says that averaging a function while walking backward along the chain in the directions according to \( S \in T_I \) approximates the conditional expectation of the function with respect to the \( \sigma \)-algebra of shift-invariant Borel sets (Fig. 3 depicts all backward walks of length 3 for \( I := \{0, 1\} \)).

5.1. **Example: the Gauss map.** Let \( X := (0, 1] \) and \( \mu \) be the Gauss measure, i.e., \( d\mu := \frac{1}{(\log 2)(1+x)} \, d\lambda \), where \( \lambda \) is Lebesgue measure. Let \( T : X \to X \) be the Gauss map \( x \mapsto \frac{1}{x} \mod 1 \). Note that identifying \( x \in (0, 1] \) with its continued fraction expression in \( \mathbb{N}^\mathbb{N} \) identifies \( T \) with the shift map \( s : \mathbb{N}^\mathbb{N} \to \mathbb{N}^\mathbb{N} \). Thus, \( T \) is countable-to-one. Moreover, \( T \) is \( \mu \)-preserving and ergodic (see [Kea95]), so Theorem 4.7 applies.

5.2. **Preliminaries on Markov chains.** Here we give definitions and basic properties of Markov chains. We will provide some short proofs to keep the paper mostly self contained, but we refer the reader to [Dur19] for a more comprehensive exposition of this subject.

We start by recalling the basic setup and notions. Let \( I \) be a countable set, which we refer to as a state space. For convenience of exposition, we will assume \( I \) is a (finite or infinite) initial segment of \( \mathbb{N} \). We equip \( I^\mathbb{N} \) with a standard Borel structure induced by the product topology, where \( I \) is discrete. In particular, the cylindrical sets

\[
[w] := \{x \in [0, 1]^I : w \subseteq x\},
\]

\( w \in I^{<\mathbb{N}} \), are clopen and form a basis for the topology. For \( w \in I^{<\mathbb{N}} \), let \( |w| \) denote the length of \( w \), and for a finite or infinite word \( w' \in I^{<\mathbb{N}} \cup I^\mathbb{N} \), let \( w \sim w' \) denote the concatenation of \( w \) and \( w' \) (i.e. the word \( w \) followed by \( w' \)).
Let $P$ be an $I \times I$ stochastic matrix, i.e. its entries are non-negative and each row adds up to 1. For each $i \in I$, we define the probability measure $\mathbb{P}_i$ supported on $[i]$ by

$$\mathbb{P}_i[\cdot^w] := P(i, w_0) \cdot P(w_1, w_2) \cdots P(w_{n-2}, w_{n-1})$$

for all $w \in I^n$, $n \in \mathbb{N}$. More generally, for any nonempty $v \in I^n$, $n \geq 1$, let $\mathbb{P}_v$ denote the probability measure supported on $[v]$ and defined by

$$\mathbb{P}_v[\cdot^w] := P(v_{n-1}, w_0) \cdot P(w_0, w_1) \cdots P(w_{m-2}, w_{m-1}),$$

for all $w \in I^m$, $m \in \mathbb{N}$. Thus, we have the following:

**Observation 5.1.** For any $i, j \in I$ and $w \in I^{<N}$,

$$\mathbb{P}_i[i^\cdot^j^w] = P(i, j)\mathbb{P}_j[w].$$

For $n \in \mathbb{N}$, let $X_n : I^N \to I$ be the random variable $x \mapsto x_n$. For $i \in I$, let $\tau_j$ be the least $n \geq 1$ such that $X_n = j$ if such exists, and $\infty$, otherwise. Further, for $i, j \in I$, let

$$\tau(i, j) := \mathbb{P}_i[\tau_j < \infty].$$

We call a state $i \in I$ **recurrent** if $\tau(i, i) = 1$, and **positive recurrent** if

$$\mathbb{E}_i \tau_i := \int_X \tau_i d\mathbb{P}_i = \sum_{1 \leq n \leq \infty} n\mathbb{P}_i[\tau_i = n] < \infty$$

(hence positive recurrent implies recurrent).

The matrix $P$ is called (positive) recurrent if such are all of its states. $P$ is said to be **irreducible** if $\tau(i, j) > 0$ for all $i, j \in I$.

**Proposition 5.2** (Recurrence is transitive). Let $P$ be a stochastic matrix on a state space $I$. For $i, j \in I$, if $i$ is recurrent and $\tau(i, j) > 0$, then $j$ is also recurrent and $\tau(j, i) = 1$. In particular, if $P$ is irreducible and has a recurrent state, then $P$ is recurrent and, in fact, $\tau(i, j) = 1$ for all $i, j \in I$.

**Proof.** This is a straightforward computation, see [Dur19, Theorem 5.3.2].

Let $\pi \in [0, 1]^I$ be a probability vector, i.e., its entries add up to 1. A (time homogeneous) Markov chain on the state space $I$ given by the transition matrix $P$ and initial distribution $\pi$ is the probability measure $\mathbb{P}$ on $I^N$ defined on the cylindrical sets $[w]$, $w \in I^{<N}$, by

$$\mathbb{P}[w] := \pi(w_0)\mathbb{P}_{w_0}[w].$$

**Assumption 5.3.** We assume throughout that all entries of $\pi$ are positive.

In light of this, observe that

$$\mathbb{P}_i = \frac{1}{\pi(i)}\mathbb{P}_i[i].$$

A probability vector $\pi \in [0, 1]^I$ is called a **stationary distribution** for $P$ if, treating it as a row-vector, we have $\pi P = \pi$. We abuse the terminology and say that a Markov chain has a property if so does its transition matrix.

The following characterizes the existence of a stationary distribution for irreducible stochastic matrices. We refer to [Dur19, Theorems 5.5.11 and 5.5.12] for a proof.

**Proposition 5.4.** If $P$ is irreducible then the following are equivalent:

1. Some state $x$ is positive recurrent.
All states are positive recurrent.

P admits a stationary distribution \( \pi \).

Furthermore, if \( P \) is irreducible and one (all) of these holds, then the stationary distribution \( \pi \) is unique and is given by \( \pi(i) = \frac{1}{E_i} \).

**Proposition 5.5.** A Markov chain is shift-invariant if and only if its initial distribution is stationary.

**Proof.** A Markov chain \( P \) is shift-invariant if and only if \( s^{-1}([j^\omega w]) \) has the same measure as \( [j^\omega w] \) for every \( j \in I \) and \( w \in I^{<N} \), i.e.

\[
P([j^\omega w]) = \sum_{i \in I} P([i^\omega j^\omega w]).
\]

But by Observation 5.1, letting \( P \) and \( \pi \) be the transition matrix and initial distribution of \( P \),

\[
\sum_{i \in I} P([i^\omega j^\omega w]) = \sum_{i \in I} \pi(i) P(i, j) P_j([j^\omega w]) = \sum_{i \in I} \pi(i) P(i, j),
\]

so \( P \) is shift-invariant if and only if for every \( j \in I \), \( \sum_{i \in I} \pi(i) P(i, j) = \pi(j) \), equivalently, \( \pi P = \pi \).

The following is the main property of Markov chains, which is usually stated in a more general form (see, for example, [Dur19, 5.2.3]), but what we state here is equivalent.

**Lemma 5.6 (Markov property).** Let \( P \) be a Markov chain. For any measurable \( A \subseteq I^N \), any \( i \in I \), and \( w \in I^{<N} \),

\[
P_{w^{-1}}(s^{-|w|}(A)) = P_i(A).
\]

**Proof.** First we check this for a basic clopen set \( A := [v] \) and we might as well assume \( v_0 = i \). Then, letting \( P \) denote the transition matrix of \( P \),

\[
P_{w^{-1}}(s^{-|w|}(A)) = P_{w^{-1}}([w^\omega v]) = P(v_0, v_1)P(v_1, v_2)\ldots P(v_{n-2}, v_{n-1}) = P_i[v] = P_i(A).
\]

This implies that the desired identity is true for all open sets since they are countable disjoint unions of basic clopen sets. This, in turn, yields the desired identity for all measurable sets by the outer regularity of \( P_{w^{-1}} \) and \( P_i \).

5.3. **Backward pointwise ergodic theorem for Markov chains.** We now give a sufficient condition for a Markov chain \( P \) to be ergodic with respect to the shift \( s \). To do so, we need auxiliary lemmas.

**Lemma 5.7 (Lebesgue density).** For any Borel measure \( \mu \) on \( I^N \) and any measurable \( A \subseteq I^N \) with \( \mu(A) > 0 \) and \( \varepsilon > 0 \), there is a positively-measured basic open set \( [w] \subseteq I^N \) with \( \frac{\mu(A \cap [w])}{\mu([w])} > 1 - \varepsilon \).
Proof. Since any Borel probability measure is outer regular, we can take an open \( U \supseteq A \) such that \( \frac{\mu(A \cap U)}{\mu(U)} > 1 - \varepsilon \). Because \( U \) is a countable disjoint union of basic open sets, there must be at least one positively-measured \([w] \subseteq U\) with \( \frac{\mu(A \cap[w])}{\mu([w])} > 1 - \varepsilon \). \( \square \)

**Lemma 5.8.** Let \( \mathbb{P} \) be a Markov chain, \( i, j \in I \), and let \( A \subseteq I^\mathbb{N} \) be a Borel set that is \( E_s \)-invariant, equivalently, \( s^{-1}(A) = A \). If \( \tau(i,j) = 1 \), then \( \mathbb{P}_i(A) = \mathbb{P}_j(A) \).

Proof. \( \tau(i,j) = 1 \) means that modulo a null set, 

\[
[i] = \bigcup_{w \in I_j \subset \mathbb{N}} [i \centerdot w] ,
\]

where \( I_j := I \setminus \{j\} \) and \( \bigcup \) denotes a disjoint union. So:

\[
\mathbb{P}_i(A) = \sum_{w \in I_j \subset \mathbb{N}} \mathbb{P}_i(A \cap [i \centerdot w]) = \sum_{w \in I_j \subset \mathbb{N}} \mathbb{P}_i(i \centerdot w - j(A)) \mathbb{P}_i[i \centerdot w - j] \\
= \mathbb{P}_j(A) \sum_{w \in I_j \subset \mathbb{N}} \mathbb{P}_i[i \centerdot w - j] = \mathbb{P}_j(A) \mathbb{P}_i[i] = \mathbb{P}_j(A).
\]

**Proposition 5.9.** If a Markov chain \( \mathbb{P} \) is irreducible and has a recurrent state, then \( \mathbb{P} \) is ergodic with respect to the shift.

Proof. Let \( A \) be an \( E_s \)-invariant set, equivalently, \( s^{-1}(A) = A \). We suppose that \( \mathbb{P}(A) > 0 \) and show that \( \mathbb{P}(A) \geq 1 - \varepsilon \), for any \( \varepsilon > 0 \). By the Lebesgue density lemma (5.7), there is a positively-measured basic open set \([w] \) such that \( \mathbb{P}_w(A) = 1 - \varepsilon \), so by the Markov property (5.6) and the shift-invariance of \( A \), \( \mathbb{P}_{w_n}(A) \geq 1 - \varepsilon \), where \( n := |w| - 1 \). By Proposition 5.2, for all \( j \in I \), \( \tau(w_n,j) = 1 \), so by Lemma 5.8, \( \mathbb{P}_j(A) = \mathbb{P}_{w_n}(A) \). This implies that \( \mathbb{P}(A) \geq 1 - \varepsilon \). \( \square \)

**Proposition 5.10.** Let \( \mathbb{P} \) be a Markov chain on a countable set \( I \) with a transition matrix \( P \) and initial distribution \( \pi \). Then the cocycle \( \rho : E_s \to \mathbb{R}^+ \) defined by \( \rho(x, s(x)) := \frac{\pi(x)}{\pi(x_1)} P(x_0, x_1) \), is the Radon–Nikodym cocycle corresponding to \( \mathbb{P} \).

Proof. To see that \( \rho \) is the Radon–Nikodym cocycle corresponding to \( \mathbb{P} \), it suffices to check that for all \( w \in I^\mathbb{N} \) and \( i \in I \), we have

\[
\mathbb{P}([i \centerdot w]) = \int_{[w]} \rho(i \centerdot x - x) \, d\mathbb{P}(x).
\]

If \( w \neq \emptyset \), then \( \rho(i \centerdot x - x) = \frac{\pi(i)}{\pi(w_0)} P(i, w_0) \) for all \( x \in [w] \). Hence,

\[
\int_{[w]} \rho(i \centerdot x - x) \, d\mathbb{P}(x) = \frac{\pi(i)}{\pi(w_0)} \cdot P(i, w_0) \cdot \mathbb{P}[w] = \mathbb{P}([i \centerdot w]).
\]
Lastly, if \( w = \emptyset \), then
\[
\int_{[w]} \rho(i^\infty, x) \, d\mathbb{P}(x) = \int_{I^n} \rho(i^\infty, x) \, d\mathbb{P}(x)
\]
\[
= \sum_{j \in I} \int_{[j]} \rho(i^\infty, x) \, d\mathbb{P}(x)
\]
\[
= \sum_{j \in I} \pi(i) P(i, j) \mathbb{P}([j])
\]
\[
= \pi(i) \sum_{j \in I} P(i, j)
\]
\[
= \pi(i) = \mathbb{P}([i]) = \mathbb{P}([i^\infty]) .
\]

Assumption 5.11. Notice that we may have \( P(i, j) = 0 \) for some \( i, j \in I \), so it may happen that \( \rho(x, y) = 0 \), i.e., \( E_s \) is not null-preserving. However, if the shift map is measure-preserving, then the set
\[
\{ x \in I^n : (\exists n) x_n = i \text{ and } x_{n+1} = j \}
\]
is null for all such pairs \( (i, j) \). Assumption 2.5 discards this set, so we may assume that \( \rho > 0 \).

Proposition 5.4, Proposition 5.5, Proposition 5.9, and Proposition 5.10 together yield the following special case of Theorem 4.7 and Corollary 4.11. Here, for any \( t \in I^n \), we put \( t \cdot x := t^\infty x \) and take \( \Gamma := I \) as our complete set of Borel partial right-inverses of \( s \).

**Corollary 5.12** (Pointwise ergodic property for Markov chains). Let \( I \) be a (countable) set, and let \( \mathbb{P} \) be an irreducible Markov chain on \( I \) with stationary initial distribution. Let \( \rho : E_n \to \mathbb{R}^+ \) be the cocycle given in Proposition 5.10. Then for any \( f \in L^1(I^n, \mathbb{P}) \), for \( \mathbb{P} \text{-a.e. } x \in I^n \), for all \( S \in \mathcal{T}_I \), \( A^{\rho}_{I^n}[S \cdot x] < \infty \), and
\[
\lim_{S \in \mathcal{T}_I} A^{\rho}_{I^n}[S \cdot x] = \int f(x) \, d\mu(x).
\]
In particular,
\[
\lim_{n \to \infty} A^{\rho}_{I^n}[\triangleright^n_s \cdot x] = \int f(x) \, d\mu(x).
\]
Furthermore, if \( f \in L^p(X, \mu), 1 \leq p < \infty \), then \( A^{\rho}_{I^n}[\triangleright^n_s \cdot x] \to_{L^p} \int f(x) \, d\mu(x) \) as \( n \to \infty \).

We now give examples of specific Markov chains satisfying the hypothesis of Corollary 5.12, that is: irreducible and with stationary initial distribution.

5.4. **Bernoulli shifts.** Perhaps the simplest example worth noting is the Bernoulli shift \( s : I^n \to I^n \), where \( I \) is finite and nonempty, and we define \( \mathbb{P} \) by \( \pi(i) := \frac{1}{|I|} \) and \( P(i, j) := \frac{1}{|I|} \) for all \( i, j \in I \). Then \( P \) irreducible since all of its entries are positive, and \( \pi P = \pi \), so by Proposition 5.4, Corollary 5.12 applies. In this case, by Proposition 5.10, the Radon–Nikodym cocycle corresponding to \( \mathbb{P} \) is given by \( \rho(x, s(x)) := \frac{1}{|I|} \) for all \( x \in I^n \).

We can also view the sequences \( x \in I^n \) as \(|I|\)-ary representations of \( x \in [0, 1] \). Thus, \( s \) is the same as the so-called baker’s map \( T : [0, 1] \to [0, 1] \) given by \( x \mapsto |I| \cdot x \mod 1 \), with Lebesgue measure \( \lambda \) on \([0, 1] \).
5.5. Boundary actions of free groups. For \( r \leq \infty \), let \( \mathbb{F}_r \) be the free group on \( r \) generators \( \{b_i\}_{i < r} \) and recall that the boundary \( \partial \mathbb{F}_r \) can be viewed as the set of all infinite reduced\(^5\) words in \( I := \{a_i\}_{i < 2r} \), where \( a_{2i} := b_i \) and \( a_{2i+1} := b_i^{-1} \) for each \( i < r \). This makes \( \partial \mathbb{F}_r \) a closed subset of \( I^\mathbb{N} \).

A class of shift-invariant measures supported on \( \partial \mathbb{F}_r \) are Markov chains \( \mathbb{P} \) on \( I \) with stationary initial distribution and transition matrix \( P \) satisfying

\[
P(a_{2i}, a_{2i+1}) = P(a_{2i+1}, a_{2i}) = 0 \quad \text{for all } i < d.
\]

The group \( \mathbb{F}_r \) naturally acts on \( \partial \mathbb{F}_r \) by concatenation: for \( w \in \mathbb{F}_r \) and \( x \in \partial \mathbb{F}_r \), \( w \cdot x := (w^x) \), where the latter denotes the reduction of the word \( w^x \). Note that this action is free on the (counitable) set of aperiodic words, where the word in \( I^\mathbb{N} \) is called periodic if it is of the form \( w^{-v}v^\infty \) for some \( w, v \in I^{<\mathbb{N}} \).

The main relevant fact about this action is that its orbit equivalence relation is the same as the orbit equivalence relation \( E_s \) induced by the shift \( s : \partial \mathbb{F}_r \to \partial \mathbb{F}_r \): indeed, for \( x \in \partial \mathbb{F}_r \), \( s(x) = x_0^{-1} \cdot x \), and conversely, for any \( a \in I \) and \( x \in I^\mathbb{N} \) with \( x_0 \neq a^{-1} \), \( a \cdot x = a^{-1} x \in s^{-1}(x) \). In particular, for each \( x \in \partial \mathbb{F}_r \), \( \triangledown_i^r(x) = B_n^{x_0} \cdot x \), where, for \( a \in I \), \( B_n^a \) is the set of all reduced words of length at most \( n \) that do not end with \( a^{-1} \).

Thus, Corollary 5.12 applied to any shift-invariant Borel probability measure \( \mu \) on \( I^\mathbb{N} \) supported on \( \partial \mathbb{F}_r \) yields a (forward) pointwise ergodic theorem for the natural action of \( \mathbb{F}_r \) on its boundary \( \partial \mathbb{F}_r \): indeed, the set \( S \cdot x \) is a set of \( (\text{forward}) \) images under the action of \( \mathbb{F}_r \) (besides being a set of preimages under the shift map \( s \)). This is the content of Theorem 1.2, which we state more precisely here:

**Theorem 5.13** (Pointwise ergodic theorem for \( \mathbb{F}_r \)). Let \( I \) be the standard symmetric set of generators of \( \mathbb{F}_r \), \( (r \leq \infty) \), and \( \mathbb{F}_r \) be the natural action of \( \mathbb{F}_r \) on its boundary \( \partial \mathbb{F}_r \subseteq I^\mathbb{N} \). Let \( \mathbb{P} \) be a Markov chain on \( I \) with stationary initial distribution and transition matrix \( P \) satisfying \( P(a, a^{-1}) = 0 \), for each \( a \in I \). Let \( E_{\mathbb{F}_r} \) denote the orbit equivalence relation induced by the action and let \( \rho : E_{\mathbb{F}_r} \to \mathbb{R}^+ \) be induced by \( \rho(i \cdot x, x) := \frac{\pi(i)}{\pi(x_0)} P(i, x) \), for \( i \in I \). For each \( f \in L^1(\partial \mathbb{F}_d, \mathbb{P}) \), for \( \mathbb{P} \)-a.e. \( x \in \partial \mathbb{F}_r \), for all \( S \in \mathcal{T}_r \), \( A_f^r(S \cdot x) < \infty \), and

\[
\lim_{S \in \mathcal{T}_r} A_f^r(S \cdot x) = \int f \, d\mathbb{P}.
\]

In particular,

\[
\lim_{n \to \infty} A_f^r(B_n^{x_0} \cdot x) = \int f \, d\mathbb{P},
\]

where \( B_n^{x_0} \subseteq \mathbb{F}_r \) denotes the set of all reduced words of length at most \( n \) that do not end with \( x_0^{-1} \). Furthermore, if \( f \in L^p(\partial \mathbb{F}_d, \mathbb{P}) \), \( 1 \leq p < \infty \), then

\[
A_f^r(B_n^{x_0} \cdot x) \to_{L_p} \int f(x) \, d\mu(x).
\]

We now give concrete examples of Markov chains satisfying the hypothesis of Theorem 5.13.

5.5.i. An example for \( r < \infty \). When \( r < \infty \), an example of such a Markov chain \( \mathbb{P} \) is the one whose initial distribution \( \pi \) is the constant vector \( \pi(a_i) := \frac{1}{2^r} \) and whose transition matrix \( P \) is defined by

\[
P(a_{2i}, a_j) := \begin{cases} 
\frac{1}{2^r-1} & \text{if } j \neq 2i + 1 \\
0 & \text{otherwise}
\end{cases}
\quad \text{and} \quad
P(a_{2i+1}, a_j) := \begin{cases} 
\frac{1}{2^r-1} & \text{if } j \neq 2i + 1 \\
0 & \text{otherwise}
\end{cases}
\]

\(^5\)A finite or infinite word \( w \) on the set \( I \) of the generators and their inverses is called reduced if a generator and its inverse do not appear side-by-side in \( w \).
for $i, j < 2r$. That is, the $i$th row of $P$ is the constant vector $\frac{1}{2r-1}$ with a zero inserted for the inverse of $a_i$. Then $P$ is irreducible since for all $i$ and $j$, if $a_i$ is not the inverse of $a_j$ then $P(a_i, a_j) > 0$, and otherwise, say if $i = 2k$ and $j = 2k + 1$, then $P(a_i, a_{j+1})P(a_{j+1}, a_j) > 0$. It is easy to check that $\pi P = \pi$, so $\pi$ is a stationary distribution, and hence $P$ satisfies the conditions of Theorem 5.13. Note that in this case, by Proposition 5.10, the Radon–Nikodym cocycle corresponding to $P$ is given by $\rho(\cdot^*x, x) := \frac{1}{2r-1}$ for all $x \in \partial F_r$ and $i \in I \setminus \{x_0^{-1}\}$.

5.5.ii. An example for $r = \infty$. We now provide an example of such a Markov chain for the action of $F_\infty$ on its boundary. Let $P$ be defined by

$$P(a_{2i}, a_j) := \begin{cases} \frac{1}{2r+1} & \text{if } j < 2i + 1 \\ 0 & \text{if } j = 2i + 1 \text{ and } P(a_{2i+1}, a_j) := \begin{cases} \frac{1}{2r+1} & \text{if } j < 2i \\ 0 & \text{if } j = 2i \\ \frac{1}{2r} & \text{if } j > 2i + 1 \end{cases} \end{cases}$$

That is, the $i$th row of $P$ is the sequence $(\frac{1}{2r+1})_{j \in \mathbb{N}}$ with a zero inserted for the inverse of $a_i$. Then $P$ is irreducible for the same reason as in the example above.

**Claim 5.14.** The state $a_0$ is positive recurrent for $P$, i.e. $\mathbb{E}_{a_0} \tau_{a_0} < \infty$.

**Proof of Claim.** We will first calculate $p_k := \mathbb{P}_{a_0}[\tau_{a_0} \geq k], k \geq 1$. To this end, for $k \geq 1$, let

$$q_k := \mathbb{P}_{a_0}[\tau_{a_0} \geq k \text{ and } X_k = a_1] \quad r_k := \mathbb{P}_{a_0}[\tau_{a_0} \geq k \text{ and } X_k \notin \{a_0, a_1\}]$$

Then

$$p_1 = 1 \quad \text{and} \quad p_{k+1} = q_k + r_k,$$

and we have the following recurrences for $k \geq 2$:

$$\begin{align*}
q_k &= \frac{1}{2}q_{k-1} + \frac{1}{4}r_{k-1}, \\
r_k &= (1 - \frac{1}{2})q_{k-1} + (1 - \frac{1}{2} - \frac{1}{4})r_{k-1} = \frac{1}{2}q_{k-1} + \frac{1}{4}r_{k-1}, \quad q_1 = 0, \\
r_1 &= 1 - \frac{1}{2} = \frac{1}{2}.
\end{align*}$$

Thus, it is clear that $q_k = r_k$ for $k \geq 2$, so $q_k = r_k = (\frac{1}{2} + \frac{1}{4})r_{k-1} = \frac{3}{4}r_{k-1} = (\frac{3}{4})^{k-2}r_2$ for $k \geq 2$. Therefore, $p_{k+1} \leq \frac{1}{2}(\frac{3}{4})^{k-2}r_2$, so $\mathbb{P}_{a_0}[\tau_{a_0} = \infty] = 0$. Furthermore, since

$$\mathbb{E}_{a_0} \tau_{a_0} \leq \sum_{1 \leq k < \infty} \frac{1}{2}p_k \leq \sum_{1 \leq k < \infty} k(\frac{3}{4})^{k-3} < \infty,$$

Claim 5.14 and Proposition 5.4 imply that $P$ is positive recurrent, i.e. $\mathbb{E}_{a_j} \tau_j < \infty$ for all $j \in \mathbb{N}$, and $\pi(a_j) := \frac{1}{\mathbb{E}_{a_j} \tau_j}$ defines a unique stationary distribution for $P$. Thus the corresponding Markov chain $P$ satisfies the hypothesis of Theorem 5.13.

6. Application to pmp actions of finitely generated groups

Let $(X, \mu)$ be a standard probability space, let $F_r$ be the free group on $r$ generators, where $r < \infty$, and let $I := \{a_i\}_{i=1}^{2r}$ be the standard symmetric set of generators. Let $\partial F_r$ be the boundary of $F_r$ (i.e., the set of all infinite reduced words in $I$), and fix a pmp action $\alpha : F_r \curvearrowright X$. Finally, let $P$ be a shift-invariant Markov measure on $\partial F_r$. We denote by $|w|$ the length of a finite word $w \in F_r$. Denote $\mathbb{P}(w) := \mathbb{P}\{x \in \partial F_r : w \subset x\}$, so

$$\mathbb{P}(w) = \begin{cases} 1 & \text{if } w = \emptyset \\ \pi(w_0)P(w_0, w_1) \cdots P(w_{n-2}, w_{n-1}) & \text{if } |w| = n > 0, \end{cases}$$
and put
\[ \mathbb{P}(S) := \sum_{w \in S} \mathbb{P}(w) \]
for a collection \( S \) of finite words.

If \( g : \mathcal{T}_I \to \mathbb{R} \), we say \( \lim_{S \in \mathcal{T}_I} g(S) = L \) if for any \( \varepsilon > 0 \), for all \( \mathbb{P} \)-large enough \( S \in \mathcal{T}_I \), \( |g(S) - L| < \varepsilon \).

We will prove the following (forward) pointwise ergodic theorem by applying Theorem 4.7 to a certain transformation of \( X \times \partial \mathbb{F}_r \).

**Theorem 6.1.** Let \( \mathbb{F}_r \) be the free group on \( r \) generators, \( r < \infty \), and let \( I \) be the standard symmetric set of generators. Let \( \mathbb{F}_r \ von X, \) be a (not necessarily free) pmp action of \( \mathbb{F}_r \).

Let \( \mathbb{P} \) be a Markov chain on \( I \) given by a stationary distribution \( \pi \) and transition matrix \( P \), where for all \( a, b \in I \), \( \pi(a) > 0 \), and \( P(a, b) = 0 \) iff \( a = b^{-1} \), this induces a probability distribution on \( I^n \) for each \( n \in \mathbb{N} \), which we still denote by \( \mathbb{P} \). Then for any \( f \in L^1(X, \mu) \), for \( \mu \)-a.e. \( x \in X \),
\[ \lim_{S \in \mathcal{T}_I} \frac{1}{\mathbb{P}(S)} \sum_{w \in S} f(w \cdot x) \mathbb{P}(w) = \overline{f}(x), \]
where \( \overline{f} \) is the conditional expectation of \( f \) with respect to the \( \sigma \)-algebra of \( G \)-invariant Borel sets.

We will ultimately obtain Theorem 6.1 by applying Theorem 4.7 to the transformation \( T : X \times \partial \mathbb{F}_r \to X \times \partial \mathbb{F}_r \), defined by
\[ (x, y) \mapsto (y_0^{-1} \cdot x, s(y)) \]
where \( s : \partial \mathbb{F}_r \to \partial \mathbb{F}_r \) is the shift map. We equip \( X \times \partial \mathbb{F}_r \) with the measure \( \mu \times \mathbb{P} \).

For \( a_{i_k} \in I \), we will also abuse notation and write \([i_0, \ldots, i_n] \) for \([a_{i_0}, \ldots, a_{i_n}] \), \( \pi(i) \) for \( \pi(a_i) \), and \( P(i_0, i_1) \) for \( P(a_{i_0}, a_{i_1}) \).

**Lemma 6.2.** \( T \) is measure-preserving.

**Proof.** Let \( U \subseteq X \) be Borel and \( i_1, \ldots, i_n < 2r \). Then for any \( x \in X \), \( i_0 < 2r \) and \( y \in [i_0] \),
\[ (x, y) \in T^{-1}(U \times [i_1, \ldots, i_n]) \iff (a_{i_0}^{-1} \cdot x, s(y)) \in U \times [i_1, \ldots, i_n] \]
\[ \iff (x, y) \in (a_{i_0} \cdot U) \times [i_0, i_1, \ldots, i_n], \]
so \( T^{-1}(U \times [i_1, \ldots, i_n]) = \bigcup_{i_0 < 2r} (a_{i_0} \cdot U) \times [i_0, i_1, \ldots, i_n] \). Hence,
\[ \nu(T^{-1}(U \times [i_1, \ldots, i_n]) = \sum_{i_0 < 2r} \nu((a_{i_0} \cdot U) \times [i_0, i_1, \ldots, i_n]) = \mu(U) \cdot \sum_{i_0 < 2r} \mathbb{P}([i_0, i_1, \ldots, i_n]) = \mu(U) \cdot \mathbb{P}(s^{-1}[i_1, \ldots, i_n]) = \nu(U \times [i_1, \ldots, i_n]). \]

**Lemma 6.3.** The Radon–Nikodym cocycle of \( E_T \) corresponding to \( \nu \) is given by
\[ \rho((x, y), T(x, y)) = \frac{\pi(y_0)}{\pi(y_1)} P(y_0, y_1). \]
Proof. It suffices to check that for Borel sets $U \subseteq X$ and $[i_0, \ldots, i_n] \subseteq \partial F$, 

$$\nu(T(U \times [i_0, \ldots, i_n])) = \int_{U \times [i_0, \ldots, i_n]} \frac{\pi(i_1)}{\pi(i_0) P(i_0, i_1)} \, d\nu.$$ 

To see this, observe that 

$$\int_{U \times [i_0, \ldots, i_n]} \frac{\pi(i_1)}{\pi(i_0) P(i_0, i_1)} \, d\nu = \frac{\pi(i_1)}{\pi(i_0) P(i_0, i_1)} \cdot \nu(U \times [i_0, \ldots, i_n])$$

$$= \frac{\pi(i_1)}{\pi(i_0) P(i_0, i_1)} \cdot \mu(U) \cdot \frac{\pi(i_0) P(i_0, i_1)}{\pi(i_1)} \cdot \mathbb{P}[i_1, \ldots, i_n]$$

$$= \mu(a \cdot U) \cdot \mathbb{P}[i_1, \ldots, i_n]$$

$$= \nu(T(U \times [i_0, \ldots, i_n])). \quad \square$$

Lemma 6.4. Let $G \curvearrowright G$ be the natural action of $G$ on its boundary; i.e., $g \cdot y = g \cdot [x, y]$ with reduction when $a$ is followed by $a^{-1}$ for any $a \in I$. Let $G \curvearrowright \times \partial G$ be the product action $g \cdot (x, y) = (g \cdot x, g \cdot y)$. Then the induced equivalence relation $E_{\alpha \times \beta}$ is the same as $E_T$ (the equivalence relation induced by $T$). 

Proof. Fix $(x, y) \in X \times \partial G$. Then for $a \in I$, if $a \neq y_0^{-1}$, then $T(a \cdot (x, y)) = T(a \cdot x, a \cdot y) = (x, y)$. If $a = y_0^{-1}$, then $T(x, y) = (a \cdot x, s(y)) = a \cdot (x, y)$. On the other hand, $T(x, y) = (y_0^{-1} x, s(y)) = y_0^{-1} \cdot (x, y)$. \quad \square

Lemma 6.5. If $G \curvearrowright \alpha X$ is ergodic, then $T$ is ergodic. 

Proof. By Lemma 6.4, it is enough to show $E_{\alpha \times \beta}$ is ergodic. Since $(\partial G, \mathbb{P})$ has the same null sets as $(\partial G, \mathbb{P}')$ (where $\mathbb{P}'$ is the uniform Markov measure as in Section 5.5.i), it is enough to show that $E_G$ is ergodic with respect to the uniform measure on $\partial G$. But by [GW16,Kai95], the natural boundary action of $G$ on $(\partial G, \mathbb{P}')$ is weakly mixing, so $E_G$ is ergodic. \quad \square

Lemma 6.6. For $f \in L^1(X, \mu)$, define $F \in L^1(X \times \partial G, \mu \times \mathbb{P})$ by $F(x, y) := f(x)$. Then for $\mu$-a.e. $x \in X$ and $\mathbb{P}$-a.e. $y \in \partial G$, $\mathbb{F}(x, y) = \mathbb{F}(x)$, where $\mathbb{F}$ is the conditional expectation of $f$ with respect to the $\sigma$-algebra of $G$-invariant Borel sets, and $\mathbb{F}$ is the conditional expectation of $F$ with respect to the $\sigma$-algebra of $T$-invariant Borel sets. 

Proof. If the action $G \curvearrowleft \alpha (X, \mu)$ is ergodic, then so is $T$ by Lemma 6.5, and hence $\mathbb{F}(x, y) = \int_{X \times \partial G} F \, d(\mu \times \mathbb{P}) = \int_X f(x) \, d\mu(x) = \overline{f}(x)$. 

In the general case, we use the Ergodic Decomposition theorem for pmp countable Borel equivalence relations (see Definition 2.6 and the following paragraph), which gives us an $E_\alpha$-ergodic decomposition $\epsilon : X \to \operatorname{Erg}_{E_\alpha}(X)$. By Proposition 2.7, for $\mu$-a.e. $x \in X$, $\overline{f}(x) = \int_X f \, d\epsilon_x$. 

But $\int_X f \, d\epsilon_x = \int_{X \times \partial G} F \, d(\epsilon_x \times \mathbb{P})$ by the definition of $F$, so it remains to show that $\mathbb{F}(x, y) = \int_{X \times \partial G} F \, d(\epsilon_x \times \mathbb{P})$ for $\mu$-a.e. $x \in X$ and $\mathbb{P}$-a.e. $y \in \partial G$. 

28
To this end, for any $H \in L^1(X \times \partial G, \mu \times \mathbb{P})$, Fubini’s theorem gives

$$\int_{X \times \partial G} H \, d(\mu \times \mathbb{P}) = \int_X \int_{\partial G} H(x,y) \, d\mathbb{P}(y) \, d\mu(x)$$

[by Proposition 2.7 with $B := X$]

$$\int_X \int_Y \int_{\partial G} H(z,y) \, d\mathbb{P}(y) \, d\epsilon_x(z) \, d\mu(x)$$

[Fubini]

$$\int_X \int_{X \times \partial G} H(z,y) \, d(\epsilon_x \times \mathbb{P})(z,y) \, d\mu(x)$$

[dummy integration]

$$\int_{X \times \partial G} \int_{X \times \partial G} H(z,y) \, d(\epsilon_x \times \mathbb{P})(z,y) \, d(\mu \times \mathbb{P})(x,y').$$

Recalling, in addition, that each measure $\epsilon_x \times \mathbb{P}$ is $T$-ergodic, by Lemma 6.5, we see that the map $(x, y) \mapsto \epsilon_x \times \mathbb{P}$ is the $E_T$-ergodic decomposition of $\mu \times \mathbb{P}$ (Definition 2.6). Thus, again by Proposition 2.7, $F(x,y) = \int_{X \times \partial G} F \, d(\epsilon_x \times \mathbb{P})$ for $(\mu \times \mathbb{P})$-a.e. $(x, y) \in X \times \partial G$, finishing the proof.

We now show that Theorem 4.7 applied to $T$ yields Theorem 6.1:

Proof. Notice that for $(x, y) \in X \times \partial G$, $T^{-1}(x, y) = \{(a_i \cdot x, a_i^{-1} y) : i < r$ and $a_i \neq y_i^{-1}\}$. Hence, the set of generators $I \subseteq G$ is a complete set of Borel partial right-inverses of $T$.

Fix $f \in L^1(X, \mu)$. Define $F \in L^1(X \times \partial G, \nu)$ by $F(x, y) := f(x)$. For $S \in \mathcal{T}_I$, and $(x, y) \in X \times \partial G$, we will abuse notation and write $|S \cdot y|^\rho_p$ for $|S \cdot (x, y)|^{(x,y)}_\rho$, as well as $\rho(y_1, y_2)$ for $\rho((x_1, y_1), (x_2, y_2))$ since $\rho$ does not depend on $x$. We may also assume without loss of generality that all sets $S$ contain only reduced words (since non-reduced words do not contribute to the weight of $S$).

Claim. For any $S \in \mathcal{T}_I$, $x \in X$, and $y_i \in [i]$ (for each $i < r$),

$$\sum_{w \in S} f(w \cdot x) \mathbb{P}(w) = \sum_{i < r} \pi(i) \sum_{w \in S, \ w_{|w|-1} \neq i^{-1}} f(w \cdot x) \rho(w^{-i}, y_i).$$

Proof of Claim. We have:

$$\sum_{i < r} \pi(i) \sum_{w \in S, \ w_{|w|-1} \neq i^{-1}} f(w \cdot x) \rho(w^{-i}, y_i) = \sum_{i < r} \pi(i) \left( f(x) + \sum_{\emptyset \neq w \notin S} f(w \cdot x) \frac{\mathbb{P}(w^{-i})}{\pi(i)} \right)$$

$$= \sum_{i < 2r} \pi(i) f(x) + \sum_{i < 2r} \sum_{\emptyset \neq w \notin S} f(w \cdot x) \mathbb{P}(w^{-i})$$

[Fubini]

$$= f(x) + \sum_{\emptyset \neq w \notin S} f(w \cdot x) \sum_{i < 2r} \mathbb{P}(w^{-i}) = f(x) + \sum_{\emptyset \neq w \notin S} f(w \cdot x) \mathbb{P}(w) = \sum_{w \in S} f(w \cdot x) \mathbb{P}(w). \quad \Box$$
Note that in the case where \( f \equiv 1 \), this claim implies \( \mathbb{P}(S) = \sum_{i < 2^r} \pi(i)|S \cdot y_i|_\rho^y \). Using this claim, we see that for any \( S \in \mathcal{T}_I \), \( x \in X \), and \( y_i \in [i] \),

\[
\frac{1}{\mathbb{P}(S)} \sum_{w \in S} f(w \cdot x)\mathbb{P}(w) = \frac{1}{\mathbb{P}(S)} \sum_{i < 2^r} \pi(i) \sum_{w \in S, \ u_i \neq w_{w^{-1}i^{-1}}} f(w \cdot x)\rho(w^{-1}y_i, y_i) = \frac{1}{\mathbb{P}(S)} \sum_{i < 2^r} \pi(i) \sum_{w \in S, \ u_i \neq w_{w^{-1}i^{-1}}} F(w \cdot x, w^{-1}y_i)\rho(w^{-1}y_i, y_i)
\]

\[
= \frac{1}{\mathbb{P}(S)} \sum_{i < 2^r} A^p_F[S \cdot (x, y_i)]\pi(i)|S \cdot y_i|_\rho^y
\]

[Claim with \( f \equiv 1 \) = \sum_{i < 2^r} A^p_F[S \cdot (x, y_i)]\pi(i)|S \cdot y_i|_\rho^y \sum_{j < 2^r} \pi(j)|S \cdot y_j|_\rho^y]

We have that Theorem 4.7 applies to \( \nu \)-a.e. \( (x, y) \in X \times \partial G \). Hence, for \( \mu \)-a.e. \( x \in X \), for all \( i < 2^r \), there is \( y_i \in \partial G \cap [i] \) such that Theorem 4.7 applies. For one of these \( x \), fix such a \( y_i \in \partial G \cap [i] \) for each \( i < 2^r \).

It now suffices to show that for any \( \epsilon > 0 \), for \( \mathbb{P} \) large enough \( S \), we have

\[
\sum_{i < 2^r} A^p_F[S \cdot (x, y_i)]\pi(i)|S \cdot y_i|_\rho^y \sum_{j < 2^r} \pi(j)|S \cdot y_j|_\rho^y \approx \epsilon \sum_{i < 2^r} F(x, y_i) \pi(i)|S \cdot y_i|_\rho^y \sum_{j < 2^r} \pi(j)|S \cdot y_j|_\rho^y;
\]

i.e., the two quantities are within \( \epsilon \), because then by (6.7), we have

\[
\frac{1}{\mathbb{P}(S)} \sum_{w \in S} f(w \cdot x)\mathbb{P}(w) \approx \sum_{i < 2^r} F(x, y_i) \frac{\pi(i)|S \cdot y_i|_\rho^y}{\sum_{j < 2^r} \pi(j)|S \cdot y_j|_\rho^y} = \sum_{i < 2^r} \mathcal{F}(x, y_i) \frac{\pi(i)|S \cdot y_i|_\rho^y}{\sum_{j < 2^r} \pi(j)|S \cdot y_j|_\rho^y} = \mathcal{F}(x).
\]

To this end, using Theorem 4.7 and Corollary 4.10, let \( M > \|\mathcal{F}(x)\| \) be large enough so that for each \( i < 2^r \), for all \( S(x, y_i) \in \mathcal{T}(x, y_i) \), \( |A^p_F[S(x, y_i)]| \leq M \), and if \( |S(x, y_i)|_\rho^y \geq M \) then \( A^p_F[S(x, y_i)] \approx \frac{\epsilon}{M} F(x, y_i) = \mathcal{F}(x) \). Let \( \delta > 0 \) be such that \( \delta < \min\{\mathbb{P}(a_i, a_j), \pi(a_k) : i, j, k < 2^r, a_i \neq a_j^{-1}\} \). Note that if \( S \) is such that for all \( i < 2^r \), \( |S \cdot (x, y_i)|_\rho^y \geq M \), then we are done.

Take \( S \in \mathcal{T}_I \) such that \( \mathbb{P}(S) > \frac{8r^2M^2}{\varepsilon^2d^2} + 1 \); i.e., \( \mathbb{P}(S \setminus \{\emptyset\}) > \frac{8r^2M^2}{\varepsilon^2d^2} \). For each \( i < 2^r \), set \( S_i := \{w \in S : w_{|w|^{-1}} = i\} \), so that \( S \setminus \{\emptyset\} = \bigsqcup_{i < 2^r} S_i \). Then by the pigeonhole principle, for some \( k < 2^r \), we have \( \mathbb{P}(S_k) > \frac{4rM^2}{\varepsilon^2d} \). Now,

\[
\text{If } a_i \neq a_k^{-1}, |S_k \cdot (x, y_i)|_\rho^y = \frac{P(k, i)}{\pi(i)} \mathbb{P}(S_k) \geq \delta \left( \frac{4rM^2}{\varepsilon^2d^2} \right) = \frac{4rM^2}{\varepsilon d} > M.
\]

We may assume that for \( a_i = a_k^{-1} \), \( |S \cdot (x, y_i)|_\rho^y < M \), since otherwise we have that for all \( i < 2^r \), \( |S \cdot (x, y_i)| \geq M \).
We claim that for each \( i < 2r \), \( A_F^\rho[S \cdot (x, y_i)] \frac{\pi(i)|S \cdot y_i|^\rho}{\sum_{j<2r} \pi(j)|S \cdot y_j|^\rho} \approx \frac{\pi(i)|S \cdot y_i|^\rho}{\sum_{j<2r} \pi(j)|S \cdot y_j|^\rho} \). First, for \( a_i = a_k^{-1} \), we have
\[
\left| A_F^\rho[S \cdot (x, y_i)] \frac{\pi(i)|S \cdot y_i|^\rho}{\sum_{j<2r} \pi(j)|S \cdot y_j|^\rho} - \frac{\pi(i)|S \cdot y_i|^\rho}{\sum_{j<2r} \pi(j)|S \cdot y_j|^\rho} \right| \leq \frac{M}{(4r\varepsilon)^2} \cdot M = \varepsilon \frac{M}{4r},
\]
so by the triangle inequality, \( A_F^\rho[S \cdot (x, y_i)] \frac{\pi(i)|S \cdot y_i|^\rho}{\sum_{j<2r} \pi(j)|S \cdot y_j|^\rho} \approx \frac{\pi(i)|S \cdot y_i|^\rho}{\sum_{j<2r} \pi(j)|S \cdot y_j|^\rho} \). On the other hand, if \( a_i \neq a_k^{-1} \), we have that \( A_F^\rho[S \cdot (x, y_i)] \approx \frac{\pi(i)|S \cdot y_i|^\rho}{\sum_{j<2r} \pi(j)|S \cdot y_j|^\rho} \), so
\[
A_F^\rho[S \cdot (x, y_i)] \frac{\pi(i)|S \cdot y_i|^\rho}{\sum_{j<2r} \pi(j)|S \cdot y_j|^\rho} \approx \frac{\pi(i)|S \cdot y_i|^\rho}{\sum_{j<2r} \pi(j)|S \cdot y_j|^\rho}.
\]
This completes the proof of the claim, and the result follows. 

\[\square\]

**References**

[Bir31] G. D. Birkhoff, *Proof of the ergodic theorem*, Proc. Natl. Acad. Sci. **17** (1931), no. 12, 656–660.

[BZ19] J. Boretsky and J. Zomback, *A tiling property for actions of amenable groups along Tempelman Følner sequences*, ArXiv e-prints (2019), available at https://arxiv.org/abs/1904.10031.

[BN13] L. Bowen and A. Nevo, *Geometric covering arguments and ergodic theorems for free groups*, Enseign. Math. (2) **59** (2013), no. 1-2, 133–164.

[BN15] L. Bowen and A. Nevo, *Von Neumann and Birkhoff ergodic theorems for negatively curved groups*, Geom. Funct. Anal. **25** (2015), no. 7, 1427–1443.

[Bu00] A. I. Bufetov, *Operator ergodic theorems for actions of free semigroups and groups*, Funktsional. Anal. i Prilozhen. **34** (2000), no. 4, 1–17, 96.

[Bu02] A. I. Bufetov, *Convergence of spherical averages for actions of free groups*, Ann. of Math. (2) **155** (2002), no. 3, 929–944.

[BK12] A. Bufetov and A. Klimenko, *On Markov operators and ergodic theorems for group actions*, European J. Combin. **33** (2012), no. 7, 1427–1443.

[But00] Sergey Butkevich, *Convergence of averages in ergodic theory*, 2000. Thesis (Ph.D.)–The Ohio State University.

[Dit92] A. Ditzen, *Definable equivalence relations on Polish spaces*, 1992. Thesis (Ph.D.)–California Institute of Technology.

[DJK94] R. Dougherty, S. Jackson, and A. S. Kechris, *The structure of hyperfinite Borel equivalence relations*, Trans. Amer. Math. Soc. **341** (1994), 193–225.

[DJ16] A. H. Dooley and K. Jarrett, *Non-singular \( \mathbb{Z}^d \)-actions: an ergodic theorem over rectangles with application to the critical dimensions*, ArXiv e-prints (2016), available at https://arxiv.org/abs/1606.01620.

[Dur19] Rick Durrett, *Probability—theory and examples*, Cambridge Series in Statistical and Probabilistic Mathematics, vol. 49, Cambridge University Press, Cambridge, 2019. Fifth edition of [MR1068527].

[FJ01a] Aihua Fan and Yunping Jiang, *On Ruelle-Perron-Frobenius operators. I. Ruelle theorem*, Comm. Math. Phys. **223** (2001), no. 1, 125–141.

[FJ01b] Aihua Fan and Yunping Jiang, *On Ruelle-Perron-Frobenius operators. II. Convergence speeds*, Comm. Math. Phys. **223** (2001), no. 1, 143–159.

[Far62] R. H. Farrell, *Representation of invariant measures*, Illinois J. Math. **6** (1962), 447–467.

[Fel07] J. Feldman, *A ratio ergodic theorem for commuting, conservative, invertible transformations with quasi-invariant measure summed over symmetric hypercubes*, Ergodic Theory Dynam. Systems **27** (2007), no. 4, 1135–1142.

[FM77] J. Feldman and C. C. Moore, *Ergodic equivalence relations and von Neumann algebras*, I. Trans. Amer. Math. Soc. **234** (1977), 289–324.
