L'Enseignement Mathématique (2) 59 (2013), 133–164

GEOMETRIC COVERING ARGUMENTS AND ERGODIC THEOREMS FOR FREE GROUPS

by Lewis BOWEN *) and Amos NEVO ‡)

ABSTRACT. We present a new approach to the proof of ergodic theorems for actions of non-amenable groups, and give here a complete self-contained account of it in the case of free groups. Our approach is based on direct geometric covering arguments and asymptotic invariance arguments generalizing those developed in the ergodic theory of amenable groups. The results we describe go beyond those previously established for measure-preserving actions of free groups, and demonstrate the significant role the boundary action of the free group plays in the ergodic theory of its measure-preserving actions. Furthermore, our approach suggests the possibility of putting the ergodic theory of amenable groups and non-amenable groups on an equal footing: both can be viewed as special cases in the general ergodic theory of amenable ergodic equivalence relations with finite invariant measure.

CONTENTS

*) supported in part by NSF grant DMS-0968762, NSF CAREER Award DMS-0954606 and BSF grant 2008274

‡) supported in part by ISF 776-09 grant and BSF grant 2008274

1. Introduction

Let Γ be a countable group and let B_t for $t \in \mathbf{N}$ or $t \in \mathbf{R}$ be a family of finite subsets of Γ. Let μ_t be a probability measure supported on B_t. Suppose Γ acts by measure-preserving transformations on a probability space (X, λ). For any $f \in L^1(X, \lambda)$ we consider the *averaging operator*

$$\mu_t(f)(x) := \sum_{\gamma \in B_t} f(\gamma^{-1}x)\mu_t(\gamma).$$

Let $\mathbf{E}[f|\Gamma]$ denote the conditional expectation of f with respect to the σ-algebra of Γ-invariant subsets. We say that $\{\mu_t\}$ is a *pointwise ergodic family in L^p* if $\mu_t(f)$ converges to $\mathbf{E}[f|\Gamma]$ pointwise almost everywhere and in L^p-norm for every $f \in L^p(X, \lambda)$ and for every measure-preserving action of Γ on a probability space (X, λ).

Some of the most useful pointwise ergodic families are those in which the sets B_t are naturally connected with the geometry of the group. A basic case to consider is when B_t is the ball of radius $t > 0$ with respect to an invariant metric, and μ_t is the uniform probability measure on B_t. Such averages are referred to as *ball averages*. Spherical averages and shell averages are defined similarly.

Most of the research on ergodic theorems has focused on the case when the group is amenable and the averages are uniformly distributed on sets which form an asymptotically invariant (Følner) sequence. The covering properties of translates of these sets and their property of asymptotic invariance play an indispensable role in the arguments developed in the amenable case. We note that ball averages on amenable groups do form an asymptotically invariant sequence in some cases, but often do not, and we refer to [Ne05] for a detailed survey of these methods and current results.

In contrast, non-amenable groups do not admit asymptotically invariant sequences, and so the arguments developed to handle amenable groups are not directly applicable. An alternative general approach to the ergodic theory of group actions based on the spectral theory of unitary representations was developed and applied first to the case where G is a semisimple S-algebraic group, and then to the case of lattice subgroups $\Gamma \subset G$. We refer to [GN10] for a detailed account of this theory. Naturally, reliance on harmonic analysis techniques limits the scope of this theory to groups whose unitary representation theory can be explicated, and to their lattice subgroups.

For general groups, and certainly for discrete groups such as (non-elementary) word-hyperbolic groups for example, spectral information is usually unavailable and harmonic analysis techniques are usually inapplicable. Exceptions do exist, and for example it was proven by spectral methods that ball averages with respect to certain invariant metrics on the free group do indeed form pointwise ergodic sequences. The metrics allowed are those arising from first fixing an embedding of the free group as a lattice in a locally compact group G. Thus in [Ne94] [NS94] the free group is viewed as a lattice in the group of automorphisms of a regular tree, and in [GN10] as a lattice in $PSL_2(\mathbf{R})$, and the metric is obtained by restricting a suitable G-invariant metric to the lattice subgroup. Note that in the case of the tree metric a periodicity phenomenon arises, and as a result the balls form a pointwise ergodic sequence if and only if the sign character of the free group does not appear in the spectrum.

In the case of spheres and balls with respect to the tree metric on the free group, a proof of the ergodic theorem in $L\log L$ was given by [Bu02], using Markov operators techniques (inspired by earlier related ideas in [Gr99]), which rely on Rota's theorem [Rot62]. This method extends to a certain extent to some groups with a Markov presentation, and in particular, to Gromov-hyperbolic groups. Thus [BKK11] and [PS] are devoted to the study of uniform averages of spherical averages on Gromov-hyperbolic groups with respect to a word metric. Norm convergence is established for integrable functions, and pointwise convergence is established for bounded functions. However, the limit function is not identified in these results, but has recently been shown to coincide with the ergodic mean in the case of surface groups in [BS10].

In the present paper we develop a new approach to proving pointwise ergodic theorems for measure-preserving actions of groups. Our method is based on intrinsic geometric covering arguments and asymptotic invariance arguments, and is completely self-contained. Our goal in what follows is to explain this general method in detail in the most accessible case, namely that of free groups, and show how to use it to generalize the existing ergodic theorems for free groups described briefly above. The main results stated below establish new maximal inequalities and pointwise convergence for a wide class of geometrically defined averages. We also establish the integrability of the maximal function associated with these sequences when the original function is in $L\log L$, and thus also pointwise convergence of the averages acting on functions in this space.

We note that our method has two significant advantages: first, it constitutes a direct generalization of the classical arguments employed to prove ergodic theorems for amenable groups, and in fact reduces the proof of ergodic theorems for the free group to the proof of ergodic theorems for a certain amenable equivalence relation. Second, as will be shown in forthcoming work, this new approach extends well beyond the class of free groups and also well beyond the specific problem of establishing pointwise ergodic theorems. For further explanation of the scope of this approach we refer to the remarks at the end of §3 below.

1.1 Statement of the main theorems

Let $\mathbf{F} = \langle a_1, \dots, a_r \rangle$ denote the free group on r generators. Let $\mathrm{S} = \{a_i, a_i^{-1}\}_{i=1}^r$ be the associated symmetric generating set. For every nonidentity element $g \in \mathbf{F}$, there is a unique sequence $t_1, \dots, t_n$ of elements in S such that $g = t_1 \cdots t_n$ and $n \geq 1$ is as small as possible. Define $|g| = n$, with $|e| = 0$. Let $\partial\mathbf{F}$ be the boundary of $\mathbf{F}$ which we identify with the set of all infinite sequences $(s_1, s_2, \dots) \in \mathrm{S}^{\mathbf{N}}$ such that $s_{i+1} \neq s_i^{-1}$ for all $i \geq 1$. If $g = t_1 \cdots t_n$ as above then the *shadow of* g (with light source at e) is the compact open set

$$O(g) = \left\{(s_1, s_2, \dots) \in \partial\mathbf{F} : \ s_i = t_i \ \text{ for } 1 \leq i \leq n\right\}.$$

The boundary admits a natural probability measure ν such that $\nu(O(g)) = (2r)^{-1}(2r-1)^{-|g|+1}$.

We denote the sphere of radius n in $\mathbf{F}$ by $S_n(e) = \{g \in \mathbf{F} : \ |g| = n\}$. Let ψ be any probability density function on $\partial\mathbf{F}$; namely $\psi \geq 0$ and $\int_{\partial\mathbf{F}} \psi \, d\nu = 1$. Define the associated probability measures μ_n^ψ on $S_n(e)$ given by $\mu_n^\psi(g) = \int_{O(g)} \psi \, d\nu$.

Let $\mathbf{F}^2 < \mathbf{F}$ be the subgroup generated by all elements g such that $|g|$ is even. It is a subgroup of index 2 in $\mathbf{F}$. Given a probability space (X, λ) on which $\mathbf{F}$ acts by measure-preserving transformations, we let $\mathbf{E}[f|\mathbf{F}^2]$ denote the conditional expectation of a function $f \in L^1(X, \lambda)$ on the σ-algebra of $\mathbf{F}^2$-invariant sets.

THEOREM 1.1. *Fix any continuous probability density function ψ on the boundary $\partial\mathbf{F}$. Then for every measure-preserving action of $\mathbf{F}$ on a standard probability space (X, λ), and for every $f \in L^p(X)$ for $1 < p < \infty$, the averages $\mu_{2n}^{\psi}(f) \in L^p(X)$ defined by*

$$\mu_{2n}^{\psi}(f)(x) := \sum_{g \in S_{2n}(e)} f(g^{-1}x)\mu_{2n}^{\psi}(g)$$

converge pointwise almost surely and in L^p-norm to $\mathbf{E}[f|\mathbf{F}^2]$. Furthermore, pointwise convergence to the same limit holds for any f in the Orlicz space $(L\log L)(X, \lambda)$.

REMARK 1.2. In the special case in which the density is identically 1, each μ_{2n} is the uniform average on $S_{2n}(e)$, and the theorem states that even-radius spherical averages converge pointwise a.e. to $\mathbf{E}[f|\mathbf{F}^2]$, for all $f \in L^p$, $1 < p < \infty$ and $f \in L\log L$. The proof of Theorem 1.1 is completely different and independent of the previous proofs of this fact in [Ne94], [NS94] and [Bu02].

REMARK 1.3. Given $w \in \mathbf{F}$, define the probability density $\rho_w = \chi_{O_w}/\nu(O_w)$ to be the normalized characteristic function of the basic compact open subset O_w of $\partial\mathbf{F}$. Thus, the sequence $\mu_{2n}^{\rho_w}$ of uniform averages on the set of all words of length $2n \geq |w|$ with initial subword w is a pointwise ergodic sequence. It is natural to call these averages (in analogy with the hyperbolic plane) *sector averages.*

Theorem 1.1 is a special case of a more general result, whose statement requires further notation. For $g \in \mathbf{F}$, let $\delta_g \in \ell^1(\mathbf{F})$ be the function $\delta_g(g') = 1$ if $g = g'$ and 0 otherwise. Let $\pi_\partial\colon \ell^1(\mathbf{F}) \to L^1(\partial\mathbf{F}, \nu)$ be the linear map satisfying $\pi_\partial(\delta_g) = \nu(O(g))^{-1}\chi_{O(g)}$ where $\chi_{O(g)}$ is the characteristic function of $O(g)$. Note that if $\mu \in \ell^1(\mathbf{F})$ and $\mu \geq 0$ then $\pi_\partial(\mu) \geq 0$ and $\|\pi_\partial(\mu)\|_1 = \|\mu\|_1$. Thus if μ_{2n} is a sequence of probability measures on $\mathbf{F}_r$, and $\pi_\partial(\mu_{2n})$ converges in $L^q(\partial\mathbf{F}_r, \nu)$ to some limit function ψ, then ψ is necessarily a probability density.

THEOREM 1.4. *Let $\{\mu_{2n}\}_{n=1}^{\infty}$ be a sequence of probability measures in $\ell^1(\mathbf{F})$ such that μ_{2n} is supported on the sphere $S_{2n}(e)$. Let $1 < q < \infty$, and suppose $\{\pi_\partial(\mu_{2n})\}_{n=1}^{\infty}$ converges in $L^q(\partial\mathbf{F}, \nu)$. Let (X, λ) be a probability space on which $\mathbf{F}$ acts by measure-preserving transformations. If $f \in L^p(X)$, $1 < p < \infty$ and $\frac{1}{p} + \frac{1}{q} < 1$, then the sequence $\{\mu_{2n}(f)\}_{n=1}^{\infty} \subset L^p(X)$ defined by*

$$\mu_{2n}(f)(x) := \sum_{g \in S_{2n}(e)} f(g^{-1}x)\mu_{2n}(g)$$

converges pointwise almost surely and in L^p-norm to $\mathbf{E}[f|\mathbf{F}^2]$. Furthermore, if $q = \infty$ and $\{\pi_\partial(\mu_{2n})\}_{n=1}^{\infty}$ converges uniformly, then pointwise convergence to the same limit holds for any f in the Orlicz space $(L\log L)(X, \lambda)$.

Theorem 1.1 follows from Theorem 1.4. To see this, fix a continuous probability density ψ on $\partial\mathbf{F}$. Then the associated averages μ_n^ψ (defined above) satisfy $\lim_{n\to\infty} \pi_\partial(\mu_n^\psi) = \psi$ in $L^q(\partial\mathbf{F}, \nu)$, for all $1 \le q \le \infty$. Indeed, the continuous functions $\pi_\partial(\mu_n^\psi)$ converge to ψ uniformly. Hence Theorem 1.4 applies.

Theorem 1.4 demonstrates the fundamental role that the boundary $\partial\mathbf{F}$ plays in the ergodic theory of $\mathbf{F}$, and raises several natural questions. For example, does the conclusion of Theorem 1.4 hold if the hypothesis that μ_{2n} is supported on $\{g \in \mathbf{F} : |g| = 2n\}$ is weakened to the condition $\lim_{n\to\infty} \mu_{2n}(g) = 0$ for all $g \in \mathbf{F}$? Does it hold if the inequality $\frac{1}{p} + \frac{1}{q} < 1$ is replaced by the weaker constraint $\frac{1}{p} + \frac{1}{q} \le 1$? What if instead of being convergent in $L^q(\partial\mathbf{F}, \nu)$, $\{\pi_\partial(\mu_{2n})\}_{n=1}^{\infty}$ is only required to be pre-compact or norm-bounded?

Finally, we remark that motivated by the results of the present paper, in the recent preprint [BK] the authors give another proof of Theorem 1.4, using the Markov operators method developed in [Bu02], which relies on Rota's theorem [Rot62] and a $0-2$ law for Markov operators.

1.2 On the ideas behind the proof

To illustrate our approach, consider the following scenario. Suppose that G is a group and $H < G$ is a subgroup. We say that H has the *automatic ergodicity property* if whenever G acts on a probability space (X, μ) by measure-preserving transformations ergodically then the action restricted to H is also ergodic. In this case, any pointwise ergodic sequence for H is a pointwise ergodic sequence for G. If H is amenable then we can use the classical theory of amenable groups to find such an ergodic sequence supported in H. Then conjugate copies of such an ergodic sequence can be averaged to construct additional pointwise ergodic sequences, giving rise to geometric averages supported on G.

For example, if $G = SL_2(\mathbf{R})$ then, by the Howe-Moore Theorem, any closed noncompact subgroup $H < G$ has the automatic ergodicity property. Thus the foregoing observation can be used to prove pointwise ergodic theorems for $SL_2(\mathbf{R})$ actions by averaging on conjugates of a horospherical (unipotent) subgroup, which is isomorphic to $\mathbf{R}$ in this case. Similar considerations apply to other Lie groups as well.

To handle free groups we will have to modify this approach by considering an appropriately chosen "amenable measurable subgroup". This "subgroup" is a sub-equivalence relation $\mathcal{R}$ of the orbit relation of $\mathbf{F}$ acting on its boundary, which we call the *horospherical sub-relation*. Whenever $\mathbf{F}$ acts on a probability space $\mathbf{F} \curvearrowright (X, \lambda)$, there is a natural extension $\mathbf{F} \curvearrowright (X \times \partial\mathbf{F}, \lambda \times \nu)$ and a natural sub-equivalence relation $\mathcal{R}^X$ of the orbit relation of $\mathbf{F}$ acting on $X \times \partial\mathbf{F}$. We show that if the action $\mathbf{F} \curvearrowright (X, \lambda)$ is ergodic then this sub-relation has at most 2 ergodic components. Moreover, this sub-relation is amenable, and in fact, hyperfinite. Theorem 1.4 is obtained by first averaging over finite-sub-equivalence relations of this sub-relation and proving their convergence by a direct geometric covering argument, and then averaging the result over the boundary $\partial\mathbf{F}$. This method establishes convergence for a variety of sequences on $\mathbf{F}$, since in the final argument we can average with respect to a variety of probability densities (or measures) on the boundary. For a more detailed overall description of the proof we refer to §4.

1.3 Outline of the paper

We begin by proving ergodic theorems for hyperfinite equivalence relations in §2. This involves a direct generalization of classical arguments. In §3 we review the boundary of $\mathbf{F}$. In §4.1 we turn to ergodicity and periodicity, and prove that an ergodic action of $\mathbf{F}$ gives rise to a 'virtually ergodic' sub-equivalence relation. In the last section we integrate the averages on the sub-relation over the boundary and prove Theorem 1.4.

2. An ergodic theorem for equivalence relations

2.1 Ergodic equivalence relations

Let (B, ν) be a standard Borel probability space. A Borel equivalence relation $\mathcal{R} \subset B \times B$ is called *discrete* if every equivalence class is finite or countable. It is called *finite* if every equivalence class is finite. Let c denote counting measure on B (so $c(E) = \#E$, $\forall E \subset B$). The measure ν on B is $\mathcal{R}$*-invariant* if $\nu \times c$ restricted to $\mathcal{R}$ equals $c \times \nu$ restricted to $\mathcal{R}$. A Borel map $\phi: B \to B$ is an *inner automorphism* of $\mathcal{R}$ if it is invertible with Borel inverse and its graph is contained in $\mathcal{R}$. Let $[\mathcal{R}]$ denote the group of inner automorphisms, also called the *full group*. If ν is $\mathcal{R}$-invariant then $\phi_*\nu = \nu$ for every $\phi \in [\mathcal{R}]$. For the rest of this section, we assume ν is an $\mathcal{R}$-invariant Borel probability measure on B.

A basic example to keep in mind is the following special case: suppose G is a discrete group acting by measure-preserving transformations on (B, ν). Then the orbit-equivalence relation $\mathcal{R} := \{(b, gb) : b \in B, g \in G\}$ is such that ν is $\mathcal{R}$-invariant. In fact, a result of [FM77] implies that all probability measure-preserving discrete equivalence relations arise from this construction (up to isomorphism).

The relation $\mathcal{R}$ is *hyperfinite* if there is an increasing sequence $\{\mathcal{R}_n\}_{n=1}^{\infty}$ of finite Borel sub-equivalence relations (namely equivalence relations whose equivalence classes are finite) whose union is $\mathcal{R}$. For $b \in B$, let $\mathcal{R}_n(b)$ denote the $\mathcal{R}_n$-equivalence class of b. For $f \in L^1(B)$, let

$$\mathbf{A}[f|\mathcal{R}_n](b) := \frac{1}{|\mathcal{R}_n(b)|} \sum_{b' \in \mathcal{R}_n(b)} f(b');$$

$$\mathbf{A}'[f|\mathcal{R}_n](b) := \frac{1}{|\mathcal{R}_n(b) \setminus \mathcal{R}_{n-1}(b)|} \sum_{b' \in \mathcal{R}_n(b) \setminus \mathcal{R}_{n-1}(b)} f(b').$$

We are interested in the convergence properties of these averages, which play the roles of balls and spheres of radius n in the $\mathcal{R}$-equivalence class of b. To describe the limit function we recall the following two definitions. A set $E \subset B$ is $\mathcal{R}$*-invariant* if $(E \times B) \cap \mathcal{R} = (B \times E) \cap \mathcal{R} = (E \times E) \cap \mathcal{R}$ (up to $\nu \times c$-measure zero). For a Borel function f on B, let $\mathbf{E}[f|\mathcal{R}]$ denote the conditional expectation of f with respect to the σ-algebra of $\mathcal{R}$-invariant Borel sets and the measure ν.

2.2 An ergodic theorem for hyperfinite relations

The main result of this section is the following pointwise convergence theorem.

THEOREM 2.1. *Let $\mathcal{R}$ be a hyperfinite Borel equivalence relation on (B, ν). Assume ν is $\mathcal{R}$-invariant. Let $\{\mathcal{R}_n\}_{n=1}^{\infty}$ be an increasing sequence of finite subequivalence relations whose union is $\mathcal{R}$. Then for any $f \in L^1(B)$,*

$\mathbf{A}[f|\mathcal{R}_n]$ converges pointwise a.e. to $\mathbf{E}[f|\mathcal{R}]$, and in the L^1-norm. Moreover, if there is a constant $C > 0$ such that

$$C|\mathcal{R}_n(b) \setminus \mathcal{R}_{n-1}(b)| \geq |\mathcal{R}_n(b)| \tag{2.1}$$

for a.e. b and every n then $\mathbf{A}'[f|\mathcal{R}_n]$ also converges pointwise a.e. and in L^1-norm to $\mathbf{E}[f|\mathcal{R}]$, as $n \to \infty$.

Theorem 2.1 is obtained from the next two theorems which are also proven in this section. Before stating them, define $L^1_0(B)$ to be the set of all functions $f \in L^1(B)$ with $\mathbf{E}[f|\mathcal{R}] = 0$ a.e.

THEOREM 2.2 (Dense set of good functions). *With the hypotheses as in Theorem 2.1, there exists a dense set $\mathcal{G} \subset L^1(B)$ such that for all $f \in \mathcal{G}$, $\mathbf{A}[f|\mathcal{R}_n]$ and $\mathbf{A}'[f|\mathcal{R}_n]$ converge pointwise a.e. to $\mathbf{E}[f|\mathcal{R}]$. Similarly, there exists a dense set $\mathcal{G}_0 \subset L^1_0(B)$ with same property.*

Consider now the maximal functions

$$\mathbf{M}[f] := \sup_n \mathbf{A}[|f||\mathcal{R}_n], \qquad \mathbf{M}'[f] := \sup_n \mathbf{A}'[|f||\mathcal{R}_n].$$

THEOREM 2.3 (weak-type L^1 maximal inequality). *With the hypotheses as in Theorem 2.1, for every $f \in L^1(B)$ and any $t > 0$,*

$$\nu\left(\{b \in B : \mathbf{M}[f](b) > t\}\right) \leq \frac{\|f\|_1}{t}.$$

Moreover, if (2.1) holds for a.e. b and every n then

$$\nu\left(\{b \in B : \ \mathbf{M}'[f](b) > t\}\right) \leq \frac{C\|f\|_1}{t}.$$

Given the two foregoing results, Theorem 2.1 follows easily using the standard classical argument.

Proof of Theorem 2.1. Let $f \in L^1(B)$. We will show that $\{\mathbf{A}[f|\mathcal{R}_n]\}_{n=1}^{\infty}$ converges pointwise a.e. $\mathbf{E}[f|\mathcal{R}]$. After replacing f with $f - \mathbf{E}[f|\mathcal{R}]$ if necessary we may assume that $\mathbf{E}[f|\mathcal{R}] = 0$ a.e.

For $\delta > 0$, let $E_\delta := \{b \in B : \ \limsup_{n\to\infty} |\mathbf{A}[f|\mathcal{R}_n](b)| \leq \delta\}$. We will show that each E_δ has measure one. Let $\epsilon = \frac{\delta^2}{4}$. According to Theorem 2.2, there exists a function $f_1 \in L^1(B)$ with $\|f - f_1\|_1 < \epsilon$ such that $\{\mathbf{A}[f_1|\mathcal{R}_n]\}_{n=1}^{\infty}$

converges pointwise a.e. to 0. Let $n > 0$. Observe that

$$|\mathbf{A}[f|\mathcal{R}_n]| \leq |\mathbf{A}[f-f_1|\mathcal{R}_n]| + |\mathbf{A}[f_1|\mathcal{R}_n]| \leq \mathbf{M}[f-f_1] + |\mathbf{A}[f_1|\mathcal{R}_n]| .$$

Let

$$D := \{b \in B : \ \mathbf{M}[f-f_1](b) \leq \sqrt{\epsilon}\} .$$

Since $\mathbf{A}[f_1|\mathcal{R}_n]$ converges pointwise a.e. to zero, for a.e. $b \in D$ there is an $N > 0$ such that $n > N$ implies

$$|\mathbf{A}[f|\mathcal{R}_n](b)| \leq \mathbf{M}[f-f_1](b) + |\mathbf{A}[f_1|\mathcal{R}_n](b)| \leq 2\sqrt{\epsilon} = \delta .$$

Hence $D \subset E_\delta$ (up to a set of measure zero). By Theorem 2.3,

$$\nu(E_\delta) \geq \nu(D) \geq 1 - \epsilon^{-1/2}\|f-f_1\|_1 > 1 - \sqrt{\epsilon} = 1 - \frac{\delta}{2} .$$

For any $\delta_1 < \delta_2$, $E_{\delta_1} \subset E_{\delta_2}$. So $\nu(E_{\delta_2}) \geq \nu(E_{\delta_1}) \geq 1 - \frac{\delta_1}{2}$ for all $\delta_1 < \delta_2$ which implies $\nu(E_{\delta_2}) = 1$. So the set $E := \cap_{n=1}^\infty E_{1/n}$ has full measure. This implies pointwise convergence of $\{\mathbf{A}[f|\mathcal{R}_n]\}_{n=1}^\infty$.

The fact that $\mathbf{A}[f|\mathcal{R}_n]$ converges to $\mathbf{E}[f|\mathcal{R}]$ in $L^1(B)$ follows from the pointwise result. To see this, observe that it is true if $f \in L^\infty(B)$ by the bounded convergence theorem. Since L^∞ is dense in L^1 and $\mathbf{A}[f|\mathcal{R}_n]$ is a contraction in L^1 this implies the result. The proof for $\mathbf{A}'$ in place of $\mathbf{A}$ is similar. □

2.3 Asymptotic invariance and pointwise convergence

Theorem 2.2 is based on the following asymptotic invariance argument. Before stating it, recall that $[\mathcal{R}_m]$ denotes the full group of the equivalence relation $\mathcal{R}_m$, so that each automorphism $\phi \in [\mathcal{R}_m]$ acts as a permutation when restricted to the finite equivalence class $\mathcal{R}_m(b)$, for almost every $b \in B$. Since $\mathcal{R}_m \subset \mathcal{R}_n$ when $n \geq m$, the equivalence class $\mathcal{R}_n(b)$ is a union of $\mathcal{R}_m$-equivalence classes, and hence $\phi(\mathcal{R}_n(b)) = \mathcal{R}_n(b)$ for almost all $b \in B$.

Lemma 2.4. *Let $\phi \in [\mathcal{R}_m]$ for some $m > 0$ and let $f \in L^\infty(B)$. Then $\mathbf{A}[f - f\circ\phi|\mathcal{R}_n]$ converges pointwise a.e. to $\mathbf{E}[f - f\circ\phi|\mathcal{R}] = 0$ as $n \to \infty$. Similarly, $\mathbf{A}'[f - f\circ\phi|\mathcal{R}_n]$ converges pointwise a.e. to 0 as $n \to \infty$.*

Proof. For any $b \in B$,

$$\begin{aligned}\lim_{n\to\infty} \left|\mathbf{A}[f - f\circ\phi|\mathcal{R}_n](b)\right| &= \lim_{n\to\infty}\left|\frac{1}{|\mathcal{R}_n(b)|}\sum_{b'\in\mathcal{R}_n(b)} f(b') - f(\phi(b'))\right| \\ &\leq 2\|f\|_\infty \lim_{n\to\infty}\frac{|\mathcal{R}_n(b)\Delta\phi(\mathcal{R}_n(b))|}{|\mathcal{R}_n(b)|} = 0 .\end{aligned}$$

The last equation holds because if $n \geq m$ then $\mathcal{R}_n(b) = \phi(\mathcal{R}_n(b))$ (since the $\mathcal{R}_n$-equivalence class of b is preserved by ϕ). Note also that if $n > m$ then

$$\mathcal{R}_n(b) \setminus \mathcal{R}_{n-1}(b) = \phi(\mathcal{R}_n(b)) \setminus \phi(\mathcal{R}_{n-1}(b)) = \phi(\mathcal{R}_n(b) \setminus \mathcal{R}_{n-1}(b)).$$

So the same argument shows that $\lim_{n\to\infty} \left|\mathbf{A}'[f - f \circ \phi|\mathcal{R}_n](b)\right| = 0$.

Since ν is $\mathcal{R}$-invariant, $\mathbf{E}[f|\mathcal{R}] = \mathbf{E}[f \circ \phi|\mathcal{R}]$. Hence $\mathbf{E}[f - f \circ \phi|\mathcal{R}] = 0$ a.e. So this proves the lemma. □

Proof of Theorem 2.2. Let $\mathcal{I} \subset L^2(B)$ be the space of $\mathcal{R}$-invariant L^2 functions. That is, $f \in \mathcal{I}$ if and only if $f(b) = f(b')$ for a.e. $(b, b') \in \mathcal{R}$. Let $\mathcal{G}_0 \subset L^2(B)$ be the space of all functions of the form $f - f \circ \phi$ for $f \in L^\infty(B)$ and $\phi \in [\mathcal{R}_m]$ for some $m > 0$. We claim that the span of $\mathcal{I}$ and $\mathcal{G}_0$ is dense in $L^2(B)$. To see this, let f_* be a function in the orthocomplement of $\mathcal{G}_0$. Denoting the L^2 inner product by $\langle \cdot, \cdot \rangle$, we have

$$0 = \langle f_*, f - f\circ\phi \rangle = \langle f_*, f \rangle - \langle f_*, f\circ\phi \rangle = \langle f_*, f \rangle - \langle f_*\circ\phi^{-1}, f \rangle = \langle f_* - f_*\circ\phi^{-1}, f \rangle$$

for any $f \in L^\infty(B)$ and $\phi \in \cup_{m=1}^\infty [\mathcal{R}_m]$. Since $L^\infty(B)$ is dense in $L^2(B)$, we have $f_* = f_* \circ \phi^{-1}$ for all $\phi \in \cup_{m=1}^\infty [\mathcal{R}_m]$. Because $\cup_{m=1}^\infty \mathcal{R}_m = \mathcal{R}$, this implies $f_* \in \mathcal{I}$; i.e., $f(b) = f(b')$ for almost every $(b, b') \in \mathcal{R}$. Since f_* is arbitrary, this implies $\mathcal{I}$ and $\mathcal{G}_0$ span $L^2(B)$ as claimed.

By Lemma 2.4 for every $f \in \mathcal{I} + \mathcal{G}_0$, $\mathbf{A}[f|\mathcal{R}_n]$ and $\mathbf{A}'[f|\mathcal{R}_n]$ converge pointwise a.e. to $\mathbf{E}[f|\mathcal{R}]$. Since $\mathcal{I} + \mathcal{G}_0$ is dense in $L^2(B)$, which is dense in $L^1(B)$, the first statement of the theorem follows. The second is similar. □

2.4 Covering argument and weak-type maximal inequality

Proof of Theorem 2.3. For $n > 0$, let

$$\mathbf{M}_n[f](b) := \max_{1\leq i\leq n} \mathbf{A}[|f||\mathcal{R}_i](b).$$

Let $D_{n,t} := \{b \in B : \mathbf{M}_n[f](b) > t\}$. We will show that $\nu(D_{n,t}) \leq \frac{\|f\|_1}{t}$ for each $n > 0$.

Let $\rho' \colon D_{n,t} \to \mathbf{N}$ be the function $\rho'(b) = m$ if $m \leq n$ is the smallest integer such that $\mathbf{A}[|f||\mathcal{R}_m](b) > t$. Let $\rho \colon D_{n,t} \to \mathbf{N}$ be the function $\rho(b) = k$ where $k = \rho'(b')$ is the largest number so that there exists $b' \in D_{n,t}$ with $b \in \mathcal{R}_{\rho'(b')}(b')$. Note that $\mathbf{A}[|f||\mathcal{R}_{\rho(b)}](b) > t$ for every $b \in D_{n,t}$, and so

$$\nu(D_{n,t}) \leq \frac{1}{t} \int_{D_{n,t}} \mathbf{A}[|f||\mathcal{R}_{\rho(z)}](z)\, d\nu(z).$$

Note further that for almost every $x, y \in D_{n,t}$, the sets $\mathcal{R}_{\rho(x)}(x)$ and $\mathcal{R}_{\rho(y)}(y)$ are either identical, or disjoint.

Let $K\colon B \times B \to \mathbf{R}$ be the function

$$K(y,z) = \frac{|f(y)|}{|\mathcal{R}_{\rho(z)}(z)|}$$

if $z \in D_{n,t}$ and $y \in \mathcal{R}_{\rho(z)}(z)$. Let $K(y,z) = 0$ otherwise. Since $\nu \times c|_{\mathcal{R}} = c \times \nu|_{\mathcal{R}}$,

$$\begin{aligned}\int_{D_{n,t}} |f(y)|\, d\nu(y) &= \int \sum_{z \in D_{n,t}} K(y,z)\, d\nu(y) = \int \sum_{y \in D_{n,t}} K(y,z)\, d\nu(z)\\ &= \int_{D_{n,t}} \mathbf{A}[|f||\mathcal{R}_{\rho(z)}](z)\, d\nu(z)\,.\end{aligned}$$

So

$$\nu(D_{n,t}) \leq \frac{1}{t}\int_{D_{n,t}} \mathbf{A}[|f||\mathcal{R}_{\rho(z)}](z)\, dz = \frac{1}{t}\int_{D_{n,t}} |f(y)|\, d\nu(y) \leq \frac{\|f\|_1}{t}\,.$$

Because this holds for all $n > 0$, this proves the first statement. Now suppose there is a constant $C > 0$ such that

$$C|\mathcal{R}_n(b) \setminus \mathcal{R}_{n-1}(b)| \geq |\mathcal{R}_n(b)|$$

for a.e. b and all n. Note that if $\mathbf{A}'[|f||\mathcal{R}_n](b) > t$ then $\mathbf{A}[|f||\mathcal{R}_n](b) > t/C$. Therefore,

$$\nu\big(\{b \in B :\ \mathbf{M}'[f](b) > t\}\big) \leq \nu\big(\{b \in B :\ \mathbf{M}[f](b) > t/C\}\big) \leq \frac{C\|f\|_1}{t}\,.$$

This concludes the proof of Theorem 2.3. $\square$

3. The free group and its boundary

3.1 The boundary action

Let $\mathbf{F} = \mathbf{F}_r = \langle a_1, \ldots, a_r\rangle$ be the free group of rank $r \geq 2$. Let $\mathrm{S} = \{a_i^{\pm 1} :\ 1 \leq i \leq r\}$. The *reduced form* of an element $g \in \mathbf{F}$ is the expression $g = s_1 \cdots s_n$ with $s_i \in \mathrm{S}$ and $s_{i+1} \neq s_i^{-1}$ for all i. It is unique. Define $|g| := n$, the length of the reduced form of g. The distance function on $\mathbf{F}$ is defined by $d(g_1, g_2) := |g_1^{-1} g_2|$.

The boundary of $\mathbf{F}$ is the set of all sequences $\xi = (\xi_1, \xi_2, \ldots) \in \mathrm{S}^{\mathbf{N}}$ such that $\xi_{i+1} \neq \xi_i^{-1}$ for all $i \geq 1$. We denote it by $\partial\mathbf{F}$. A metric d_∂ on $\partial\mathbf{F}$ is defined by $d_\partial\big((\xi_1, \xi_2, \ldots), (t_1, t_2, \ldots)\big) = \frac{1}{n}$ where n is the largest natural number such that $\xi_i = t_i$ for all $i < n$. If $\{g_i\}_{i=1}^{\infty}$ is any sequence of elements in $\mathbf{F}$ and $g_i := t_{i,1} \cdots t_{i,n_i}$ is the reduced form of g_i then $\lim_i g_i = (\xi_1, \xi_2, \ldots) \in \partial\mathbf{F}$ if $t_{i,j}$ is eventually equal to ξ_j for all j. If $\xi \in \partial\mathbf{F}$ then denote by $\xi_i \in \mathrm{S}$ the i-th element in the sequence $\xi = (\xi_1, \xi_2, \xi_3, \ldots)$.

We define a probability measure ν on $\partial\mathbf{F}$ as follows. For every finite sequence $t_1, \ldots, t_n$ with $t_{i+1} \neq t_i^{-1}$ for $1 \le i < n$, let

$$\nu\Big(\big\{(\xi_1, \xi_2, \ldots) \in \partial\mathbf{F} : \ \xi_i = t_i \,\forall 1 \le i \le n\big\}\Big) := |S_n(e)|^{-1} = (2r-1)^{-n+1}(2r)^{-1}.$$

By the Carathéodory extension Theorem, this uniquely extends to a Borel probability measure ν on $\partial\mathbf{F}$.

There is a natural action of $\mathbf{F}$ on $\partial\mathbf{F}$ by

$$(t_1 \cdots t_n)\xi := (t_1, \ldots, t_{n-k}, \xi_{k+1}, \xi_{k+2}, \ldots),$$

where $t_1, \ldots, t_n \in \mathrm{S}$, $t_1 \cdots t_n$ is in reduced form and k is the largest number $\le n$ such that $\xi_i^{-1} = t_{n+1-i}$ for all $i \le k$. Observe that if $g = t_1 \cdots t_n$ then the Radon-Nikodym derivative satisfies

$$\frac{d\nu \circ g}{d\nu}(\xi) = (2r-1)^{2k-n}.$$

Note that the level set of the Radon-Nikodym derivative for a given $\xi \in \partial\mathbf{F}$, namely $\big\{g \in \mathbf{F}\,;\ \frac{d\nu\circ g}{d\nu}(\xi) = 1\big\}$ consists of those words g of even length $n = 2k$ (say), whose last k letters form a word which is the inverse of the word formed by the first k letters of ξ.

3.2 The horospherical and synchronous tail relations

Let $\mathcal{R}$ be the equivalence relation on $\partial\mathbf{F}$ given by $(\xi, \eta) \in \mathcal{R}$ if and only if when writing $\xi = (\xi_1, \xi_2, \ldots)$ and $\eta = (\eta_1, \eta_2, \ldots)$, there exists n such that $\eta_i = \xi_i$ for all $i > n$. Thus $\eta\mathcal{R}\xi$ if and only if η and ξ have the same (synchronous) tail, if and only if they differ by finitely many coordinates only.

Let $\mathcal{R}_n$ be the equivalence relation given by $(\xi, \eta) \in \mathcal{R}_n$ if and only if $\xi_i = \eta_i$ for all $i > n$. Then $\mathcal{R}$ is the increasing union of the finite subequivalence relations $\mathcal{R}_n$. Thus $\mathcal{R}$ is hyperfinite.

Consider now the relation $\mathcal{R}'$ on $\partial\mathbf{F}$ such that $\eta\mathcal{R}'\xi$ if and only if there exists $g \in \mathbf{F}$ such that $g\xi = \eta$ and $\frac{d\nu\circ g}{d\nu}(\xi) = 1$. Recall that according to our description above of the words belonging to level set, g has even length $n = 2k$ (say), and the word formed by its last k letters coincides with the inverse of the word formed by the first k letters of ξ. It follows that $\eta = g\xi$ has the same synchronous tail as ξ, from the $(k+1)$-th letter onwards. Equivalently, g^{-1} belongs to the horosphere based at ξ and passing through the identity in $\mathbf{F}$, namely the geodesic from g^{-1} to ξ and the geodesic from e to ξ meet at a point (namely the word formed from the first k letters of ξ and g^{-1}) which is equidistant from e and g^{-1}. It is therefore natural to call $\mathcal{R}'$ the *horospherical relation*: $\eta\mathcal{R}'\xi$ if and only if $\eta = g\xi$, where g^{-1} belongs to the horosphere based at ξ passing through the identity. Similarly, it is natural to call the equivalence class of ξ under $\mathcal{R}_n$ the horospherical-ball of radius n based at ξ.

Since ξ and $\eta = g\xi$ have the same synchronous tail, $\mathcal{R}'$ coincides with the synchronous tail relation $\mathcal{R}$, and so it is an equivalence relation. We note that this property is in fact a consequence just of the fact that the Radon-Nikodym derivative is a cocycle and $\mathcal{R}'$ constitutes its kernel. Thus to see that $\mathcal{R}'$ is symmetric note that for $\xi = g^{-1}\eta$, we have $\frac{d\nu\circ g^{-1}}{d\nu}(\eta) = \left(\frac{d\nu\circ g}{d\nu}(\xi)\right)^{-1} = 1$, so that the relation is indeed symmetric. The transitivity of the horospherical relation $\mathcal{R}'$ follows from the cocycle identity which the Radon-Nikodym derivative satisfies. Finally note that by definition, the measure ν is $\mathcal{R}'$-invariant.

3.3 On the scope of the method

Let us make three brief remarks on the scope of our approach.

Remark 3.1. The observations in the previous subsection point to the following underlying fundamental idea. Utilizing the action of a discrete group on a suitable boundary B and the associated Radon-Nikodym derivative, it is possible to define a useful notion of "horospheres" and "horoballs" in the group, using the level sets of the Radon-Nikodym derivatives. Furthermore, B carries an associated equivalence relation, which is amenable and has an invariant probability measure. Finally, there exist natural subsets in the equivalence classes which are asymptotically invariant under the equivalence relation in a suitable sense.

Given a measure-preserving action on X, the equivalence relation on B can be extended to product space $X \times B$, which again has an invariant probability measure and asymptotically invariant subsets on the equivalence classes. We may integrate the averages on these sets over B and thus construct averages on the group itself, and then deduce pointwise ergodic theorems for the group action on X from a pointwise ergodic theorem for the equivalence relation. This general approach to the problem is developed in [BN1] and [BN4], and more concretely for Gromov-hyperbolic groups in [BN3].

REMARK 3.2. It is natural to consider the synchronous tail relation, more generally, in any Markov chain coding the elements of a Markov group. We can then appeal to the general pointwise convergence result we have established for hyperfinite equivalence relations with finite invariant measure in Theorem 2.1. This approach to proving ergodic theorems for Markov groups is developed in [BN5].

REMARK 3.3. The geometric covering arguments we have used in the proof of Theorem 2.1 can of course be replaced by an appeal to the martingale convergence theorem. Nevertheless, we have chosen to give a self-contained proof, the reason being that this proof can be greatly generalized. In particular, it leads to a proof of a ratio ergodic theorem for non-singular hyperfinite equivalence relations in the absence of an invariant measure, where the martingale theorem does not apply. In turn, this implies a horospherical ratio ergodic theorem for non-singular actions of the free groups, and other Markov groups. This approach to the ratio ergodic theorem for non-amenable groups is developed in [BN2].

4. IDENTIFYING THE LIMIT IN THE ERGODIC THEOREM

The present section and the following one are devoted to the proof of Theorem 1.4, and we begin with a very brief description of our proof plan. Given an ergodic probability preserving action of $\mathbf{F}$ on (X, μ), we will consider $X \times \partial\mathbf{F}$ with the measure $\mu \times \nu$. We extend the horospherical relation $\mathcal{R}$ on $\partial\mathbf{F}$ to a relation $\mathcal{R}^X$ on $X \times \partial\mathbf{F}$, which is still hyperfinite, with invariant measure $\mu \times \nu$. We consider the operators of averaging on the finite classes approximating $\mathcal{R}^X$, and appeal to Theorem 2.1, which guarantees the averages converge pointwise to the conditional expectation on the σ-algebra of $\mathcal{R}^X$-invariant sets. We then prove the crucial fact that this σ-algebra coincides with the σ-algebra of subsets of $X \times \partial\mathbf{F}$ invariant under the action of $\mathbf{F}^2$, the subgroup of $\mathbf{F}$ consisting of words of even length ("automatic ergodicity"). We then use the fact that the action of $\mathbf{F}^2$ on $\partial\mathbf{F}$ is *weak-mixing*, namely that the product action of $\mathbf{F}^2$ on $X \times \partial\mathbf{F}$ is ergodic for every probability measure preserving ergodic action of $\mathbf{F}^2$ on a space X. These arguments identify the limit in the ergodic theorem we seek to prove, as the conditional expectation on the σ-algebra of $\mathbf{F}^2$-invariant sets. We will then integrate the averages defined on the finite classes in $X \times \partial\mathbf{F}$, over the boundary $\partial\mathbf{F}$. This gives rise to averages acting on functions on X, given by certain probability measures on $\mathbf{F}$, which converge to the limit stated in Theorem 1.4. Analyzing these measures, we recognize that they coincide with the measures described in Theorem 1.4.

4.1 Automatic ergodicity of the extended horospherical relation

Let $\mathbf{F}$ act on a standard probability space (X, λ) by measure-preserving transformations. Let $\mathbf{F}^2$ be the subgroup of $\mathbf{F}$ generated by words of length 2, which has index 2 in $\mathbf{F}$. For any $f \in L^1(X)$, let $\mathbf{E}[f|\mathbf{F}^2] \in L^1(X)$ be the conditional expectation of f on the σ-algebra of $\mathbf{F}^2$-invariant measurable sets.

Let $\mathcal{R}^X$ be the equivalence relation on $X \times \partial\mathbf{F}$ given by $((x,\xi),(x',\xi')) \in \mathcal{R}^X \Leftrightarrow \exists g \in \mathbf{F}$ with $x = gx'$, $\xi = g\xi'$ and $(\xi,\xi') \in \mathcal{R}$. Equivalently, $\frac{d\nu\circ g}{d\nu}(\xi') = 1$, or ξ and ξ' have the same synchronous tail and g^{-1} is in the horosphere based at ξ' passing through e. For $f \in L^1(X \times \partial\mathbf{F})$, let $\mathbf{E}[f|\mathcal{R}^X]$ denote the conditional expectation of f on the sigma-algebra of $\mathcal{R}^X$-invariant sets.

For $f \in L^1(X)$, define $i(f) \in L^1(X \times \partial\mathbf{F})$ by $i(f)(x,\xi) = f(x)$. The map $f \mapsto i(f)$ isometrically embeds in $L^1(X)$ into $L^1(X \times \partial\mathbf{F})$. The purpose of this section is to prove the following "automatic ergodicity" property:

Theorem 4.1. *For any* $f \in L^1(X)$, $\mathbf{E}\left[i(f)|\mathcal{R}^X\right] = i(\mathbf{E}[f|\mathbf{F}^2])$.

Similar results were proven in [Bo08] for all word hyperbolic groups.

We remark that it is necessary to consider the action of $\mathbf{F}^2$ rather than $\mathbf{F}$. For example, if X is a two-point set, λ is the uniform probability measure and all generators $\{a_1,\dots,a_r\}$ of $\mathbf{F}$ act nontrivially on X then the action of $\mathbf{F}$ on X is ergodic but the equivalence relation $\mathcal{R}^X$ on $X \times \partial\mathbf{F}$ is not.

Theorem 4.1 is based on a more general result. Before stating it, we introduce the following definitions.

Definition 4.2. Let $\mathcal{I} \subset L^1(X \times \partial\mathbf{F})$ be the σ-algebra of sets A that are invariant under the relation $\mathcal{R}^X$, i.e., for all $\phi \in [\mathcal{R}^X]$ (the full group of $\mathcal{R}^X$) we have $\phi(A) = A$.

Let $\mathcal{I}_2 \subset L^1(X \times \partial\mathbf{F})$ be the σ-algebra of $\mathbf{F}^2$-invariant sets A, namely such that for all $g \in \mathbf{F}^2$, $g(x,\xi) = (gx, g\xi) \in A$ if and only if $(x,\xi) \in A$.

The corresponding conditional expectations are denoted by $\mathbf{E}[F|\mathcal{R}^X]$ and $\mathbf{E}[F|\mathcal{I}_2]$ (for $F \in L^1(X \times \partial\mathbf{F})$).

In the next subsection we prove:

THEOREM 4.3. *For any* $F \in L^1(X \times \partial\mathbf{F})$, $\mathbf{E}[F|\mathcal{I}_2] = \mathbf{E}[F|\mathcal{R}^X]$, *or equivalently* $\mathcal{I} = \mathcal{I}_2$.

Theorem 4.1 clearly follows from Theorem 4.3 and the following lemma, whose proof is included below.

LEMMA 4.4. *For any* $f \in L^1(X)$, $\mathbf{E}[i(f)|\mathcal{I}_2] = i\big(\mathbf{E}[f|\mathbf{F}^2]\big)$.

Proof. Without loss of generality, we may assume that the action of $\mathbf{F}^2$ on (X, λ) is ergodic. It suffices to show that the diagonal action $\mathbf{F}^2 \curvearrowright X \times \partial\mathbf{F}$ is ergodic.

Let μ be the uniform measure on the generating set S. Then the action of $\mathbf{F}$ on the Poisson boundary of the random walk determined by μ is canonically identified with the action of $\mathbf{F}$ on $(\partial\mathbf{F}, \nu)$ (e.g., see [Ka00]). Note that the support of the convolution $\mu * \mu$ generates $\mathbf{F}^2$. Hence the action of $\mathbf{F}^2$ on the Poisson boundary of the random walk determined by $\mu * \mu$ is identified with the action of $\mathbf{F}^2$ on $(\partial\mathbf{F}, \nu)$. By [Ka03] and [AL05], this action is weakly mixing. This implies the diagonal action of $\mathbf{F}^2$ on $(\partial\mathbf{F} \times X, \nu \times \lambda)$ is ergodic. □

4.2 IDENTIFYING THE LIMIT: PROOF OF AUTOMATIC ERGODICITY

We now turn to the proof of Theorem 4.3.

DEFINITION 4.5. For $(x, \xi) \in X \times \partial\mathbf{F}$, write $\xi = (\xi_1, \xi_2, \ldots)$, and define $\mathrm{P}(x, \xi) \in X \times \partial\mathbf{F}$ by $\mathrm{P}(x, \xi) = \xi_1^{-1}(x, \xi)$. More generally, if $n \geq 1$ then let $\mathrm{P}^n(x, \xi) := (\xi_1 \cdots \xi_n)^{-1}(x, \xi)$.

LEMMA 4.6. *Let* $F \in L^1(X \times \partial\mathbf{F})$. *If* $F \circ \mathrm{P}^2 = F$ *a.e. then* F *is measurable w.r.t.* $\mathcal{I}_2$, *namely* F *is* $\mathbf{F}^2$*-invariant.*

Proof. Let $(x, \xi) \in X \times \partial\mathbf{F}$ and $g = t_1 \cdots t_{2n} \in \mathbf{F}^2$ be in reduced form. By definition,

$$g\xi = (t_1, \ldots, t_{2n-k}, \xi_{k+1}, \xi_{k+2}, \ldots),$$

where k is the largest number such that $\xi_i^{-1} = t_{2n+1-i}$ for all $i \leq k$. For any $x \in X$, if k is even then $(gx, g\xi) \in \mathrm{P}^{-(2n-k)}\mathrm{P}^k(x, \xi)$. If k is odd then $(gx, g\xi) \in \mathrm{P}^{-(2n-k+1)}\mathrm{P}^{k+1}(x, \xi)$. Thus if $f \circ \mathrm{P}^2 = f$ a.e. then $f \circ g = f$ a.e.. This implies the lemma. □

PROPOSITION 4.7. *To prove Theorem* 4.3, *it suffices to prove that* $F \circ \mathrm{P}^2 = F$ *for all* F *measurable w.r.t.* $\mathcal{I}$, *namely for all* $\mathcal{R}^X$*-invariant functions.*

Proof. From Lemma 4.6 it follows that if $F \circ \mathrm{P}^2 = F$ for all $\mathcal{R}^X$-invariant functions F then $\mathcal{I} \subset \mathcal{I}_2$. To see the reverse inclusion, let $(x, \xi), (x', \xi') \in X \times \partial\mathbf{F}$ be $\mathcal{R}^X$-equivalent. By definition, there exists $g \in \mathbf{F}$ such that $(x', \xi') = (gx, g\xi)$. As noted in §3.2 above, g is necessarily in $\mathbf{F}^2$. Thus if F if $\mathbf{F}^2$-invariant then for a.e. pair $(x, \xi), (x', \xi')$ of $\mathcal{R}^X$-equivalent points in $X \times \partial\mathbf{F}$, $F(x, \xi) = F(x', \xi')$, namely F is $\mathcal{R}^X$-invariant. This shows $\mathcal{I}_2 \subset \mathcal{I}$. $\square$

The next proposition is the key geometric result in the proof of Theorem 4.3. Define $\mathrm{P}_\partial \colon \partial\mathbf{F} \to \partial\mathbf{F}$ by $\mathrm{P}_\partial(\xi) = \xi_1^{-1}\xi$. Recall that d_∂ is a distance function on $\partial\mathbf{F}$ defined by $d_\partial\big((\xi_1, \xi_2, \ldots), (t_1, t_2, \ldots)\big) = \frac{1}{n}$, where n is the largest natural number such that $\xi_i = t_i$ for all $i < n$.

PROPOSITION 4.8. *There exist measurable maps* $\psi_n, \omega_n \colon \partial\mathbf{F} \to \partial\mathbf{F}$ *(for* $n > 5$*) such that*

(1) $\forall \xi \in \partial\mathbf{F}$, $d_\partial(\xi, \omega_n\xi) = \frac{1}{n}$;

(2) $\forall \xi \in \partial\mathbf{F}$, $d_\partial\big(\psi_n\omega_n(\xi), \mathrm{P}_\partial^2\,\omega_n(\xi)\big) = \frac{1}{n-1}$;

(3) *the graphs of* ω_n *and* ψ_n *are contained in* $\mathcal{R}$;

(4) $\forall \xi \in \partial\mathbf{F}$, $\exists g \in \mathbf{F}$ *such that* $\psi_n\omega_n(\xi) = g\omega_n(\xi)$ *and* $\mathrm{P}_\partial^2\,\omega_n(\xi) = g\xi$. *Except for a countable set of* $\xi \in \partial\mathbf{F}$, g *is uniquely determined.*

(5) $\forall f \in L^1(\partial\mathbf{F})$, $\|f \circ \omega_n\|_1 = \|f\|_1$ *and* $\|f \circ \psi_n\|_1 \le (2r-1)^2$, *where* r *is the rank of the free group* $\mathbf{F}$.

Proof. We begin by defining ω_n and ψ_n. Recall that $\mathrm{S} = \{a_1, \ldots, a_r, a_1^{-1}, \ldots, a_r^{-1}\}$ is the chosen generating set of $\mathbf{F}$. Let $K \colon \mathrm{S}^3 \to \mathrm{S}^3$ be a bijection so that for any $(s_{k-1}, s_k, s_{k+1}) \in \mathrm{S}^3$, $K(s_{k-1}, s_k, s_{k+1}) = (s_{k-1}, s'_k, s_{k+1})$ for some $s'_k \notin \{s_{k-1}^{-1}, s_k, s_{k+1}^{-1}\}$.

We now fix $n > 5$ and define $\omega_n \colon \partial\mathbf{F} \to \partial\mathbf{F}$ by $\omega_n(s_1, s_2, \ldots) = (t_1, t_2, \ldots)$ where $t_i = s_i$ for all $i \neq n$ and $t_n = s'_n$ where $K(s_{n-1}, s_n, s_{n+1}) = (s_{n-1}, s'_n, s_{n+1})$. By its definition ω_n is invertible, Borel, $d_\partial(\xi, \omega_n(\xi)) = \frac{1}{n}$ for any $\xi \in \partial\mathbf{F}$ and $(\omega_n)_*\nu = \nu$. Moreover since ω_n does not change the tail of the sequence (i.e., because $t_i = s_i$ for all sufficiently large i), the graph of ω_n is contained in $\mathcal{R}$. Because ω_n is measure-preserving, $\|f \circ \omega_n\|_1 = \|f\|_1$ for any $f \in L^1(\partial\mathbf{F})$.

Define $\psi_n \colon \partial\mathbf{F} \to \partial\mathbf{F}$ by

$$\psi_n\omega_n(s_1, s_2, \ldots) = (s_3, \ldots, s_{n-1}, s'_n, s_n^{-1}, s'_n, s_{n+1}, s_{n+2}, \ldots),$$

where $K(s_{n-1}, s_n, s_{n+1}) = (s_{n-1}, s_n', s_{n+1})$. Because ω_n is invertible, ψ_n is well-defined.

Note that the m-th coordinate of $\psi_n\omega_n(s_1, s_2, \ldots)$ equals the m-th coordinate of $\omega_n(s_1, s_2, \ldots)$ if $m \geq n$. Therefore, the graph of ψ_n is contained in $\mathcal{R}$. If $\xi = (s_1, s_2, \ldots)$ then

$$\mathrm{P}_\partial^2\, \omega_n(\xi) = (s_3, \ldots, s_{n-1}, s_n', s_{n+1}, \ldots).$$

Thus $d_\partial(\psi_n\omega_n(\xi), \mathrm{P}_\partial^2\, \omega_n(\xi)) = \frac{1}{n-1}$. This concludes the proofs of parts (1), (2) and (3).

Note that setting $g = (s_3 \cdots s_{n-1})s_n'(s_1 \cdots s_n)^{-1}$ we have $\mathrm{P}_\partial^2\, \omega_n(\xi) = g\xi$, and similarly, $\psi_n\omega_n(\xi) = g\omega_n(\xi)$. Since only countably many points $\xi \in \partial\mathbf{F}$ have a non-trivial stabilizer in $\mathbf{F}$, and g is uniquely determined when the stabilizer of ξ is trivial, and this proves part (4). Anticipating the argument of the next proposition, let us note that $s_1 s_2 g$ satisfies $s_1 s_2 g\xi = \omega_n(\xi)$.

We now claim that ψ_n is at most $(2r-1)^2$-to-1 (that is, for each $b \in \partial\mathbf{F}$, b has at most $(2r-1)^2$-preimages under ψ_n). Because ω_n is invertible, it suffices to show that $\psi_n\omega_n$ is at most $(2r-1)^2$-to-1. Suppose that $(u_1, u_2, \ldots) \in \partial\mathbf{F}$ and

$$\psi_n\omega_n(u_1, u_2, \ldots) = \psi_n\omega_n(s_1, s_2, \ldots) = (s_3, \ldots, s_{n-1}, s_n', s_n^{-1}, s_n', s_{n+1}, s_{n+2}, \ldots).$$

By definition of $\psi_n\omega_n$, $u_i = s_i$ for $i \geq 3$. Since there are $(2r-1)^2$ choices for (u_1, u_2) the claim follows.

Since the graph of ψ_n is contained in $\mathcal{R}$ and ν is $\mathcal{R}$-invariant, the claim implies $\|f \circ \psi_n\|_1 \leq (2r-1)^2\|f\|_1$ for all $f \in L^1(\partial\mathbf{F})$, and this proves part (5). $\square$

LEMMA 4.9. *There exist measurable maps $\Phi_n, \Psi_n, \Omega_n \colon X \times \partial\mathbf{F} \to X \times \partial\mathbf{F}$ (for $n > 5$) such that*

(1) *for all $F \in L^1(X \times \partial\mathbf{F})$, $\lim_{n\to\infty} \|F \circ \Psi_n \circ \Omega_n - F \circ \mathrm{P}^2 \circ \Phi_n\|_1 = 0$;*

(2) *for all $F \in L^1(X \times \partial\mathbf{F})$, $\lim_{n\to\infty} \|F \circ \Omega_n - f\|_1 = 0$;*

(3) *the graphs of Φ_n and Ψ_n are contained in $\mathcal{R}^X$.*

Proof. For $n > 5$ an integer, let ψ_n and ω_n be as in Proposition 4.8. Fix $(x, \xi) \in X \times \partial\mathbf{F}$ and let $g_1, g_2 \in \mathbf{F}$ be such that $g_1\xi = \omega_n(\xi)$ and $g_2\xi = \psi_n(\xi)$. As noted already, for almost every ξ (and every n), g_1 and g_2 are uniquely determined. Define $\Omega_n(x, \xi) := (x, g_1\xi)$, $\Phi_n(x, \xi) := (g_1 x, g_1\xi)$ and $\Psi_n(x, \xi) := (g_2 x, g_2\xi)$.

Since the graphs of ψ_n and ω_n are contained in $\mathcal{R}$, the graphs of Φ_n and Ψ_n are contained in $\mathcal{R}^X$. Let d_X be a metric on X that induces its

Borel structure and turns X into a compact metric space. For $(x,\xi),(x',\xi')\in X\times\partial\mathbf{F}$, define $d_*((x,\xi),(x',\xi'))=d_X(x,x')+d_\partial(\xi,\xi')$. By Proposition 4.8(1), $d_*(\Omega_n(x,\xi),(x,\xi))=d_\partial(\omega_n(\xi),\xi)=1/n$. Furthermore

$$d_*(\Psi_n\circ\Omega_n(x,\xi),\mathrm{P}^2\circ\Phi_n(x,\xi))=d_*((g_2x,\psi_n\omega_n(\xi)),P^2(g_1x,\omega_n(\xi))),$$

where $g_1\xi=\omega_n(\xi)$, and $g_2\omega_n(\xi)=\psi_n\omega_n(\xi)$. Writing $\xi=(s_1,s_2,s_3,\dots)$, we have $P^2(g_1x,\omega_n(\xi))=(s_2^{-1}s_1^{-1}g_1x,P^2_\partial\omega_n(\xi))$. Recalling from the proof of Proposition 4.8(2) that $g=(s_3\cdots s_{n-1})s_n'(s_1\cdots s_n)^{-1}$ satisfies $\mathrm{P}^2_\partial\,\omega_n(\xi)=g\xi$ as well as $\psi_n\omega_n(\xi)=g\omega_n(\xi)$, we conclude that $g=g_2=s_2^{-1}s_1^{-1}g_1$. Since by Proposition 4.8(2), we have $d_\partial\big(\psi_n\omega_n(\xi),\mathrm{P}^2_\partial\,\omega_n(\xi)\big)=\frac{1}{n-1}$ it follows that $d_*(\Psi_n\circ\Omega_n(x,\xi),\mathrm{P}^2\circ\Phi_n(x,\xi))=\frac{1}{n-1}$, for almost every $(x,\xi)\in X\times\partial\mathbf{F}$.

Thus if F is a continuous function on $X\times\partial\mathbf{F}$ then the bounded convergence theorem implies

$$\lim_{n\to\infty}\|F\circ\Psi_n\circ\Omega_n-F\circ\mathrm{P}^2\circ\Phi_n\|_1=0$$
$$\lim_{n\to\infty}\|F\circ\Omega_n-F\|_1=0\,.$$

It follows from Proposition 4.8(5) that the operators $F\mapsto F\circ\Omega_n$, $F\mapsto F\circ\Phi_n$ and $F\mapsto F\circ\Psi_n$ are all bounded for $F\in L^1(X\times\partial\mathbf{F})$ with bound independent of n. It is easy to see that $F\mapsto F\circ\mathrm{P}^2$ is also a bounded operator on $L^1(X\times\partial\mathbf{F})$. Since the continuous functions are dense in $L^1(X\times\partial\mathbf{F})$, this implies the lemma. □

We can now prove Theorem 4.3.

Proof of Theorem 4.3. By Proposition 4.7, it suffices to show that $F\circ\mathrm{P}^2=F$ for every F which is $\mathcal{R}^X$-invariant. Let Φ_n,Ψ_n,Ω_n, $n>5$ be as in Lemma 4.9. Because F is $\mathcal{R}^X$-invariant and the graph of Ψ_n is contained in $\mathcal{R}^X$, it follows that $F\circ\Psi_n=F$ for all n. An easy exercise shows that P preserves the equivalence relation: if (x,ξ) is $\mathcal{R}^X$-equivalent to (y,ξ') then $\mathrm{P}(x,\xi)$ is $\mathcal{R}^X$-equivalent to $\mathrm{P}(y,\xi')$. It follows that $F\circ\mathrm{P}^2$ is also $\mathcal{R}^X$-invariant. Now since the graph of Φ_n is also contained in $\mathcal{R}^X$, it follows that $F\circ\mathrm{P}^2\circ\Phi_n=F\circ\mathrm{P}^2$ for all n. We now have

$$\begin{aligned}\|F-F\circ\mathrm{P}^2\|_1&=\|F-F\circ\mathrm{P}^2\circ\Phi_n\|_1\\&\le\|F-F\circ\Psi_n\circ\Omega_n\|_1+\|F\circ\Psi_n\circ\Omega_n-F\circ\mathrm{P}^2\circ\Phi_n\|_1\\&=\|F-F\circ\Omega_n\|_1+\|F\circ\Psi_n\circ\Omega_n-F\circ\mathrm{P}^2\circ\Phi_n\|_1\,.\end{aligned}$$

The previous lemma now implies $F=F\circ\mathrm{P}^2$ as claimed. □

5. PROOFS OF ERGODIC THEOREMS

5.1 CONVERGENCE ALONG THE HYPERFINITE RELATION ON $X \times \partial\mathbf{F}$

Let $\mathbf{F}$ act by measure-preserving transformations on a probability space (X, λ). Let $\mathcal{R}_n^X$ be the equivalence relation on $X \times \partial\mathbf{F}$ defined by $((x, \xi), (x', \xi')) \in \mathcal{R}_n^X$ if and only if there exists $g \in \mathbf{F}$ with $(gx, g\xi) = (x', \xi')$ and $(\xi, \xi') \in \mathcal{R}_n$ (i.e., if $\xi = (\xi_1, \ldots) \in \mathrm{S}^{\mathbf{N}}$ and $\xi' = (\xi_1', \ldots) \in \mathrm{S}^{\mathbf{N}}$ then $\xi_i = \xi_i'$ for all $i > n$).

We will need the next easy lemma (which is left as an exercise).

LEMMA 5.1. *For any $(x, \xi) \in X \times \partial\mathbf{F}$,*

$$|\mathcal{R}_n^X(x, \xi)| = |\mathcal{R}_n(\xi)| = (2r-1)^n .$$

So

$$\begin{aligned}|\mathcal{R}_n^X(x, \xi) \setminus \mathcal{R}_{n-1}^X(x, \xi)| &= |\mathcal{R}_n(\xi) \setminus \mathcal{R}_{n-1}(\xi)| = (2r-1)^{n-1}(2r-2) \\ &= \frac{2r-2}{2r-1}|\mathcal{R}_n(\xi)| = \frac{2r-2}{2r-1}|\mathcal{R}_n^X(x, \xi)| .\end{aligned}$$

For $f \in L^p(X, \lambda)$, recall that $i(f) \in L^p(X \times \partial\mathbf{F})$ is the function $i(f)(x, \xi) = f(x)$. Collecting results of the previous sections, we can now prove:

COROLLARY 5.2. *For $f \in L^1(X)$, let $\mathbf{E}[f|\mathbf{F}^2]$ be the conditional expectation of f with respect to the σ-algebra of $\mathbf{F}^2$-invariant sets. Then for $\lambda \times \nu$-a.e. $(x, \xi) \in X \times \partial\mathbf{F}$,*

$$\mathbf{E}[f|\mathbf{F}^2](x) = \lim_{n \to \infty} \mathbf{A}'[i(f)|\mathcal{R}_n^X](x, \xi) .$$

Proof. Lemma 5.1 implies that the assumption in Theorem 2.1 is satisfied, and thus for a.e. $(x, \xi) \in X \times \partial\mathbf{F}$,

$$\mathbf{E}[i(f)|\mathcal{R}^X](x, \xi) = \lim_{n \to \infty} \mathbf{A}'[i(f)|\mathcal{R}_n^X](x, \xi) .$$

By Theorem 4.1, $\mathbf{E}\left[i(f)|\mathcal{R}^X)\right](x, \xi) = \mathbf{E}[f|\mathbf{F}^2](x)$ for a.e. $(x, \xi) \in X \times \partial\mathbf{F}$. $\square$

In the next section, we will need the following strong L^p-maximal inequality. For $f \in L^1(X \times \partial\mathbf{F})$, define

$$\mathbf{M}'[f] := \sup_n \mathbf{A}'[|f||\mathcal{R}_n^X] .$$

PROPOSITION 5.3. *For every $p > 1$ there is a constant $C_p > 0$ such that for every $f \in L^p(X \times \partial\mathbf{F})$, $\|\mathbf{M}'[f]\|_p \le C_p \|f\|_p$. Moreover, there is a constant C_1 such that if $f \in L\log^+ L(X,\lambda)$, then*

$$\|\mathbf{M}'[f]\|_{L^1} \le C_1 \|f\|_{L\log L}.$$

Proof. It follows from Theorem 2.3 that for any $f \in L^1(X \times \partial\mathbf{F})$ the weak-type (1,1) maximal inequality holds:

$$\lambda \times \nu \left(\{(x,\xi) \in X \times \partial\mathbf{F} : \ \mathbf{M}'[f] > t\}\right) \le \frac{C\|f\|_1}{t}$$

for some constant $C > 0$. (In fact we can take $C = \frac{2r-1}{2r-2}$.)

The first part of the proposition now follows from standard interpolation arguments. Namely, since the operator $f \mapsto \mathbf{M}'[f]$ is of weak-type $(1,1)$ and is norm-bounded on L^∞, it is norm-bounded in every L^p, $1 < p < \infty$ (see e.g. [SW71, Ch. V, Thm 2.4]).

Finally, given the weak-type $(1,1)$ maximal inequality, the fact that when $f \in L\log^+ L(X,\lambda)$, the maximal function is in fact integrable and satisfies the Orlicz-norm bound is standard, see e.g. [DS, p. 678]. □

We now turn to the proof of Theorem 1.4, and show that by integrating the converging averages $\mathbf{A}'[i(f)|\mathcal{R}_n^X](x,\xi)$ over $\partial\mathbf{F}$, we obtain converging averages defined by probability measures on $\mathbf{F}$. We begin by considering integration with respect to weighted averages on the boundary.

5.2 AVERAGING OVER THE BOUNDARY $\partial\mathbf{F}$

From now on we let $1 < p, q < \infty$ be such that $\frac{1}{p} + \frac{1}{q} = 1$. Let $\psi \in L^q(\partial\mathbf{F}, \nu)$ be a probability density on the boundary, namely $\psi \ge 0$ and $\int \psi \, d\nu = 1$. The goal of the present subsection is to prove:

PROPOSITION 5.4. *For $f \in L^p(X,\lambda)$ and $n \ge 0$, define $\mathbf{A}'_\psi[f|\mathcal{R}_n] \in L^p(X)$ by*

$$\mathbf{A}'_\psi[f|\mathcal{R}_n](x) := \int_{\partial\mathbf{F}} \mathbf{A}'[i(f)|\mathcal{R}_n^X](x,\xi)\psi(\xi)\, d\nu(\xi).$$

Then $\mathbf{A}'_\psi[f|\mathcal{R}_n]$ converges pointwise a.e. to $\mathbf{E}[f|\mathbf{F}^2]$. Furthermore, if ψ is essentially bounded then the same conclusion holds for any $f \in L\log^+ L(X,\lambda)$.

The proof of Proposition 5.4 uses the following:

LEMMA 5.5. *Let p, q, ψ, f be as above and define*

$$\mathbf{M}'_\psi[f] := \sup_n \mathbf{A}'_\psi[|f||\mathcal{R}_n] .$$

Then there exists a constant $C_p > 0$ (depending only on p) such that for every $f \in L^p(X, \lambda)$

$$\|\mathbf{M}'_\psi[f]\|_p \leq C_p\|\psi\|_q\|f\|_p .$$

Furthermore, there is a constant $C_1 > 0$ such that if ψ is bounded, then for any $f \in L\log^+ L(X, \lambda)$ we have

$$\|\mathbf{M}'_\psi[f]\|_1 \leq C_1\|\psi\|_\infty\|f\|_{L\log L} .$$

Proof. Without loss of generality, let us assume $f \geq 0$. We start with the case $1 < p < \infty$. For a.e. $x \in X$,

$$\begin{aligned}
\left|\mathbf{M}'_\psi[f](x)\right|^p &= \sup_n \left|\mathbf{A}'_\psi[f|\mathcal{R}_n](x)\right|^p \\
&= \sup_n \left|\int_{\partial\mathbf{F}} \mathbf{A}'[i(f)|\mathcal{R}^X_n](x, \xi)\psi(\xi)\ d\nu(\xi)\right|^p \\
&\leq \sup_n \|\mathbf{A}'[i(f)|\mathcal{R}^X_n](x, \cdot)\|^p_{L^p(\partial\mathbf{F})}\|\psi\|^p_{L^q(\partial\mathbf{F})} .
\end{aligned}$$

The last line above is justified by Hölder's inequality. Next, we observe that for any $n \geq 1$,

$$\begin{aligned}
\int_X \sup_n \|\mathbf{A}'[i(f)|\mathcal{R}^X_n](x, \cdot)\|^p_{L^p(\partial\mathbf{F})}\ d\lambda(x) &= \int_X \sup_n \int_{\partial\mathbf{F}} |\mathbf{A}'[i(f)|\mathcal{R}^X_n](x, \xi)|^p\ d\nu(\xi)d\lambda(x) \\
&\leq \int_X \int_{\partial\mathbf{F}} \mathbf{M}'[i(f)](x, \xi)^p\ d\nu(\xi)d\lambda(x) \\
&= \|\mathbf{M}'[i(f)]\|^p_{L^p(X\times\partial\mathbf{F})} .
\end{aligned}$$

Putting this together with the previous inequality we obtain

$$\begin{aligned}
\|\mathbf{M}'_\psi[f]\|^p_{L^p(X)} &= \int_X \left|\mathbf{M}'_\psi[f](x)\right|^p\ d\lambda(x) \\
&\leq \|\mathbf{M}'[i(f)]\|^p_{L^p(X\times\partial\mathbf{F})}\|\psi\|^p_{L^q(\partial\mathbf{F})} .
\end{aligned}$$

The first part of the lemma now follows from

$$\begin{aligned}
\|\mathbf{M}'_\psi[f]\|_{L^p(X)} &\leq \|\psi\|_{L^q(\partial\mathbf{F})}\|\mathbf{M}'[i(f)]\|_{L^p(X\times\partial\mathbf{F})} \\
&\leq C_p\|\psi\|_{L^q(\partial\mathbf{F})}\|i(f)\|_{L^p(X\times\partial\mathbf{F})} = C_p\|\psi\|_{L^q(\partial\mathbf{F})}\|f\|_{L^p(X)} ,
\end{aligned}$$

where $C_p > 0$ is as in Proposition 5.3.

The second part of the lemma follows in exactly the same way, taking $f \in L\log^+ L(X)$, $p = 1$ and $q = \infty$ above. Using the integrability of the maximal function and the norm bound

$$\|\mathbf{M}'[i(f)]\|_{L^1(X\times\partial\mathbf{F})} \leq C_1 \|f\|_{(L\log L)(X)}$$

together with the boundedness of ψ, the desired estimate follows. □

Proof of Proposition 5.4. We will first prove the proposition in the special case in which $f \in L^\infty(X)$. By Corollary 5.2, $\mathbf{A}'[i(f)|\mathcal{R}_n^X]$ converges pointwise a.e. to $i\big(\mathbf{E}[f|\mathbf{F}^2]\big)$. By Lebesgue's dominated convergence theorem, this implies that for a.e. $x \in X$, $\{\mathbf{A}'_\psi[f|\mathcal{R}_n](x)\}_{n=1}^\infty$ converges to $\int \mathbf{E}[f|\mathbf{F}^2](x)\psi(\xi)\,d\nu(\xi) = \mathbf{E}[f|\mathbf{F}^2](x)$. This finishes the case in which $f \in L^\infty(X)$.

Now suppose that $f \in L^p(X)$. After replacing f with $f - \mathbf{E}[f|\mathbf{F}^2]$ if necessary, we may assume that $\mathbf{E}[f|\mathbf{F}^2] = 0$. Let $\epsilon > 0$. Let $f' \in L^\infty(X)$ be such that $\|f - f'\|_p < \epsilon$ and $\mathbf{E}[f'|\mathbf{F}^2] = 0$. Clearly:

$$|\mathbf{A}'_\psi[f|\mathcal{R}_n]| \leq |\mathbf{A}'_\psi[f - f'|\mathcal{R}_n]| + |\mathbf{A}'_\psi[f'|\mathcal{R}_n]| \leq \mathbf{M}'_\psi[f - f'] + |\mathbf{A}'_\psi[f'|\mathcal{R}_n]|\,.$$

Since $\mathbf{A}'_\psi[f'|\mathcal{R}_n] \to 0$ pointwise a.e., it follows that for a.e. $x \in X$,

$$\limsup_n |\mathbf{A}'_\psi[f|\mathcal{R}_n](x)| \leq \mathbf{M}'_\psi[f - f'](x)\,.$$

Lemma 5.5 now implies:

$$\|\limsup_n |\mathbf{A}'_\psi[f|\mathcal{R}_n]|\,\|_p \leq \|\mathbf{M}'_\psi[f - f']\|_p \leq C_p\|f - f'\|_p \leq C_p\epsilon\,.$$

Since $\epsilon > 0$ is arbitrary, it follows that $\|\limsup_n |\mathbf{A}'_\psi[f|\mathcal{R}_n]|\,\|_p = 0$. Equivalently, $\mathbf{A}'_\psi[f|\mathcal{R}_n]$ converges to 0 pointwise a.e.

The second part of the proposition follows similarly using approximation in the Orlicz norm. □

5.3 Convergence for probability measures on $\mathbf{F}$

We now turn to establish that each operator $f \mapsto \mathbf{A}'_\psi[f|\mathcal{R}_n]$ is given by averaging with respect to a probability measure η_{2n}^ψ on the group $\mathbf{F}$. We will then prove pointwise convergence of spherical averages and their generalizations.

Definition 5.6. Let $t_1 \cdots t_{2n} = g$ be the reduced form of an element $g \in \mathbf{F}^2$. Define

$$O'(g) = O(t_1 \cdots t_n) - O(t_1 \cdots t_n t_{n+1}) \subset \partial\mathbf{F}\,,$$

where $O(\cdot)$ is as defined in the introduction (namely $\xi = (\xi_1, \ldots) \in O(t_1 \cdots t_n) \Leftrightarrow \xi_i = t_i$, $1 \leq i \leq n$). Recall that $(\xi, \xi') \in \mathcal{R}_n \Leftrightarrow \xi_i = \xi'_i$ for all $i > n$. Thus $\xi \in O'(g)$ if and only if $(\xi, g^{-1}\xi) \in \mathcal{R}_n \setminus \mathcal{R}_{n-1}$.

Define $\eta_{2n}^{\psi} \in \ell^1(\mathbf{F})$ by

$$\eta_{2n}^{\psi}(g) := \frac{1}{(2r-2)(2r-1)^{n-1}} \int_{O'(g)} \psi \, d\nu$$

if $|g| = 2n$ and 0 otherwise.

LEMMA 5.7. *For any function $f \in L^p(X)$ $(p \geq 1)$, any $n \geq 0$ and any $x \in X$,*

$$\mathbf{A}'_{\psi}[f|\mathcal{R}_n](x) = \sum_{g \in \mathbf{F}} f(g^{-1}x)\eta_{2n}^{\psi}(g).$$

Proof. By definition

$$\mathbf{A}'[i(f)|\mathcal{R}_n^X](x, \xi) = \frac{1}{|\mathcal{R}_n^X(x, \xi) \setminus \mathcal{R}_{n-1}^X(x, \xi)|} \sum_{(x', \xi') \in \mathcal{R}_n^X(x, \xi) \setminus \mathcal{R}_{n-1}^X(x, \xi)} i(f)(x', \xi').$$

By Lemma 5.1, and since f depends only on x,

$$\mathbf{A}'[i(f)|\mathcal{R}_n^X](x, \xi) = (2r-1)^{-n+1}(2r-2)^{-1} \sum_g f(g^{-1}x),$$

where for each ξ the sum is over all $g \in \mathbf{F}$ such that $(\xi, g^{-1}\xi) \in \mathcal{R}_n(\xi) \setminus \mathcal{R}_{n-1}(\xi)$. Such g necessarily has length $2n$, and as noted above, the latter condition is equivalent to $\xi \in O'(g)$. Thus when integrating the expression

$$\mathbf{A}'_{\psi}[f|\mathcal{R}_n](x) = \int_{\partial \mathbf{F}} \mathbf{A}'[i(f)|\mathcal{R}_n^X](x, \xi)\psi(\xi) \, d\nu(\xi)$$

over the boundary, we integrate over the sets $O'(g)$, as g ranges over the sphere $S_{2n}(e)$. For a given x, in each such set $\mathbf{A}'[i(f)|\mathcal{R}_n^X](x, \xi)$ has the value given above (independent of ξ), and hence we obtain

$$= (2r-1)^{-n+1}(2r-2)^{-1} \sum_{g \in S_{2n}(e)} f(g^{-1}x) \int_{O'(g)} \psi \, d\nu = \sum_{g \in \mathbf{F}} f(g^{-1}x)\eta_{2n}^{\psi}(g). \quad \square$$

We can now state the following corollary, proved previously for L^p, $p > 1$ in [Ne94] [NS94] and for $L \log^+ L$ in [Bu02].

COROLLARY 5.8. *Let $p > 1$ and $f \in L^p(X)$ or more generally $f \in L\log^+ L(X)$. Then for a.e. $x \in X$,*

$$\mathbf{E}[f|\mathbf{F}^2](x) = \lim_{n\to\infty} \frac{1}{|S_{2n}(e)|} \sum_{g\in S_{2n}(e)} f(g^{-1}x).$$

Proof. Set $\psi \equiv 1$. By the previous lemma,

$$\begin{aligned}\mathbf{A}'_\psi[f|\mathcal{R}_n](x) &= \sum_{g\in\mathbf{F}} f(g^{-1}x)\eta^\psi_{2n}(g)\\ &= \frac{1}{(2r-2)(2r-1)^{n-1}} \sum_{g\in S_{2n}(e)} f(g^{-1}x)\nu(O'(g)).\end{aligned}$$

If $g = g_1 \cdots g_{2n}$ then $\xi \in O'(g)$ if and only if $\xi_i = g_i$ for all $1 \le i \le n$ and $\xi_{n+1} \neq g_{n+1}$. This implies $\nu(O'(g)) = (2r)^{-1}(2r-1)^{-n}(2r-2)$. So

$$\mathbf{A}'_\psi[f|\mathcal{R}_n](x) = (2r)^{-1}(2r-1)^{-2n+1} \sum_{g\in S_{2n}(e)} f(g^{-1}x) = \frac{1}{|S_{2n}(e)|} \sum_{g\in S_{2n}(e)} f(g^{-1}x).$$

By Proposition 5.4, $\mathbf{A}'_\psi[f|\mathcal{R}_n]$ converges pointwise a.e. to $\mathbf{E}[f|\mathbf{F}^2]$. □

5.4 BOUNDARY BEHAVIOR OF PROBABILITY MEASURES ON $\mathbf{F}$

Recall that we have defined $\pi_\partial\colon \ell^1(\mathbf{F}) \to L^1(\partial\mathbf{F}, \nu)$ to be the linear map satisfying $\pi_\partial(\delta_g) = \nu(O(g))^{-1}\chi_{O(g)}$.

Thus far, we have established that the probability measures η^ψ_{2n} on $\mathbf{F}$ (for a given probability density $\psi \in L^q(\partial\mathbf{F})$) have good convergence properties. Our goal is to establish good convergence properties for a general sequence μ_{2n} of probability measures, with μ_{2n} supported on $S_{2n}(e)$, where we assume that the functions $\pi_\partial(\mu_{2n})$ converge in $L^q(\partial\mathbf{F})$. The limit is then necessarily a probability density $\psi \in L^q(\partial\mathbf{F})$.

We now recall that given a probability density $\psi \in L^q(\partial\mathbf{F})$, μ^ψ_n denotes the probability measure on $\mathbf{F}$ given by

$$\mu^\psi_n(g) = \int_{O(g)} \psi(\xi)\, d\nu(\xi),$$

if g is in the sphere $S_n(e)$, and otherwise $\mu^\psi_n(g) = 0$.

We begin by showing that the two sequences $\pi_\partial(\mu^\psi_n)$ and $\pi_\partial(\eta^\psi_{2n})$ both converge to ψ in L^q-norm. This fact will be used in the next subsection, in a comparison argument which reduces the convergence of μ_{2n} to that of η^ψ_{2n}.

LEMMA 5.9. *Let $\psi \in L^q(\partial\mathbf{F}, \nu)$ be a probability density. Then the sequences $\{\pi_\partial(\mu^\psi_n)\}_{n=0}^\infty$ and $\{\pi_\partial(\eta^\psi_{2n})\}_{n=1}^\infty$ both converge to ψ in L^q-norm when $1 \le q < \infty$, and uniformly if ψ is continuous.*

Proof. For $n \geq 1$, let $\mathbf{E}[\psi|\Sigma_n]$ be the conditional expectation of ψ on Σ_n, the σ-algebra generated by $\{O(g)\colon\ g \in S_n(e)\}$. Thus

$$\mathbf{E}[\psi|\Sigma_n](\xi) = \frac{1}{\nu(O(g))} \int_{O(g)} \psi(\xi')\, d\nu(\xi'),$$

if $\xi \in O(g)$ with $g \in S_n(e)$. Note that $\nu(O(g)) = |S_n(e)|^{-1} = \frac{1}{(2r)(2r-1)^{n-1}}$, and that $\mathbf{E}[\psi|\Sigma_n](\xi) = \pi_\partial(\mu_n^\psi)(\xi)$.

We now claim that $\mathbf{E}[\psi|\Sigma_n]$ converges to ψ in L^q-norm as $n \to \infty$. This is clearly the case when ψ is continuous, and hence uniformly continuous, in which case $\mathbf{E}[\psi|\Sigma_n]$ in fact converges uniformly to ψ on $\partial\mathbf{F}$. In general when $\psi \in L^q$, there is a sequence of continuous functions ψ_k converging in L^q-norm to ψ, and the claim follows by an obvious approximation argument. (Alternatively, one can of course appeal to the martingale convergence theorem.)

Noting that

$$\begin{aligned}\pi_\partial(\eta_{2n}^\psi) &= \frac{|S_{2n}(e)|}{(2r-2)(2r-1)^{n-1}} \Big(|S_n(e)|^{-1}\mathbf{E}[\psi|\Sigma_n] - |S_{n+1}(e)|^{-1}\mathbf{E}(\psi|\Sigma_{n+1})\Big) \\ &= \frac{2r-1}{2r-2}\pi_\partial(\mu_n^\psi) - \frac{1}{2r-2}\pi_\partial(\mu_{n+1}^\psi),\end{aligned}$$

convergence of $\pi_\partial(\eta_{2n}^\psi)$ to ψ in L^q-norm follows immediately. When ψ is continuous then since $\mathbf{E}[\psi|\Sigma_n]$ converges uniformly to ψ, so does $\pi_\partial(\eta_{2n}^\psi)$. $\square$

The next result is not needed for the proof of the main theorem; we state and prove it since it seems interesting in its own right.

PROPOSITION 5.10. *As above, let $1 < p, q < \infty$ be such that $\frac{1}{p} + \frac{1}{q} = 1$. Let $f \in L^{p'}(X)$ for some p' with $p < p'$. For $x \in X$ and $n \geq 0$, define $f_{x,2n} \in \ell^p(\mathbf{F})$ by $f_{x,2n}(g) = f(g^{-1}x)$ if $g \in S_{2n}(e)$ and $f_{x,2n}(g) = 0$ otherwise. Let $\tilde{\pi}_\partial(f_{x,2n})\colon \partial\mathbf{F} \to \mathbf{R}$ be the function*

$$\tilde{\pi}_\partial(f_{x,2n})(\xi) := \sum_{g \in S_{2n}(e)} f_{x,2n}(g)\chi_{O(g)}(\xi).$$

Then, for a.e. $x \in X$, $\{\tilde{\pi}_\partial(f_{x,2n})\}_{n=1}^\infty$ converges to the constant function $\xi \mapsto \mathbf{E}[f|\mathbf{F}^2](x)$ in the weak topology on $L^p(\partial\mathbf{F}, \nu)$.

Proof. For $\rho \in L^p(\partial\mathbf{F}, \nu)$ and $\psi \in L^q(\partial\mathbf{F}, \nu)$, let $\langle\rho, \psi\rangle := \int \rho\psi\, d\nu$. It suffices to show that for any $\psi \in L^q(\partial\mathbf{F}, \nu)$ and a.e. $x \in X$, $\langle\tilde{\pi}_\partial(f_{x,2n}), \psi\rangle$ converges to $\mathbf{E}[f|\mathbf{F}^2](x) \int \psi\, d\nu$. By linearity, we may assume that $\psi \geq 0$ and

$\int \psi \, d\nu = 1$. Observe that

$$\langle \tilde{\pi}_\partial(f_{x,2n}), \psi \rangle = \langle \tilde{\pi}_\partial(f_{x,2n}), \pi_\partial(\eta^\psi_{2n}) \rangle + \langle \tilde{\pi}_\partial(f_{x,2n}), \psi - \pi_\partial(\eta^\psi_{2n}) \rangle . \tag{5.1}$$

It follows from Proposition 5.4 and Lemma 5.7, that for a.e. $x \in X$,

$$\langle \tilde{\pi}_\partial(f_{x,2n}), \pi_\partial(\eta^\psi_{2n}) \rangle = \mathbf{A}'_\psi[f|\mathcal{R}_n]$$

converges to $\mathbf{E}[f|\mathbf{F}^2](x)$. It follows from the previous lemma that $\psi - \pi_\partial(\eta^\psi_{2n})$ converges to zero in norm. Note that $\|\tilde{\pi}_\partial(f_{x,2n})\|_p^p$ involves the *uniform spherical* average of $|f|^p$:

$$\|\tilde{\pi}_\partial(f_{x,2n})\|_p^p = \int_{\partial\mathbf{F}} \left| \sum_{g \in S_{2n}} f(g^{-1}x)\chi_{O(g)}(\xi) \right|^p d\nu(\xi) =$$

$$= \sum_{g \in S_{2n}} \left| f(g^{-1}x) \right|^p \nu(O(g)) = |S_{2n}|^{-1} \sum_{g \in S_{2n}} \left| f(g^{-1}x) \right|^p .$$

Hence it follows from Corollary 5.8 that $\|\tilde{\pi}_\partial(f_{x,2n})\|_p^p$ converges to $\mathbf{E}\big[|f|^p|\mathbf{F}^2\big](x)$ for a.e. $x \in X$, where we also use $p < p'$ to conclude that $|f|^p \in L^{p'/p}(X)$ with $p'/p > 1$.

By Hölder's inequality,

$$\left| \left\langle \tilde{\pi}_\partial(f_{x,2n}), \psi - \pi_\partial(\eta^\psi_{2n}) \right\rangle \right| \leq \|\tilde{\pi}_\partial(f_{x,2n})\|_p \|\psi - \pi_\partial(\eta^\psi_{2n})\|_q$$

tends to zero as $n \to \infty$. Thus equation (5.1) implies the proposition. $\square$

REMARK 5.11. Typically, $\tilde{\pi}_\partial(f_{x,2n})$ does not converge to $\mathbf{E}[f|\mathbf{F}^2](x)$ in norm. To see this, observe that $\|\tilde{\pi}_\partial(f_{x,2n})\|_p$ converges to $\mathbf{E}[|f|^p|\mathbf{F}^2](x)^{1/p}$ (for a.e. $x \in X$). The norm of the constant function $\xi \mapsto \mathbf{E}[f|\mathbf{F}^2](x)$ is $|\mathbf{E}[f|\mathbf{F}^2](x)|$. Unless f is constant on the ergodic component containing x, Jensen's inequality implies $\mathbf{E}[|f|^p|\mathbf{F}^2](x)^{1/p} \neq |\mathbf{E}[f|\mathbf{F}^2](x)|$. This uses $p > 1$.

5.5 Proof of the main theorem

We now turn to the proof of Theorem 1.4, whose formulation we recall for the reader's convenience.

THEOREM 1.4. *Let $\{\mu_{2n}\}_{n=1}^\infty$ be a sequence of probability measures in $\ell^1(\mathbf{F})$ such that μ_{2n} is supported on $S_{2n}(e)$. Let $1 < q < \infty$, and suppose that $\{\pi_\partial(\mu_{2n})\}_{n=1}^\infty$ converges in $L^q(\partial\mathbf{F}, \nu)$. Let (X, λ) be a probability space on which $\mathbf{F}$ acts by measure-preserving transformations. If $f \in L^p(X)$, $1 < p < \infty$ and $\frac{1}{p} + \frac{1}{q} < 1$, then the averages*

$$\mu_{2n}(f)(x) := \sum_{g \in S_{2n}} f(g^{-1}x)\mu_{2n}(g)$$

converge pointwise almost surely and in L^p-norm to $\mathbf{E}[f|\mathbf{F}^2]$. *Furthermore, if* $q=\infty$ *and* $\pi_\partial(\mu_{2n})$ *converge uniformly, then pointwise convergence to the same limit holds for any* f *in the Orlicz space* $(L\log^+ L)(X,\lambda)$.

Proof of Theorem 1.4. To begin, we assume $1<q<\infty$. Let $p'>1$ be such that $\frac{1}{p'}+\frac{1}{q}=1$. Since $\frac{1}{p}+\frac{1}{q}<1$, it follows that $p'<p$. Let $f\in L^p(X)$. Choose a measurable version $\mathbf{E}[f|\mathbf{F}^2]$ of the conditional expectation. Let $\psi\in L^q(\partial\mathbf{F})$ be the probability density which is the limit of $\{\pi_\partial(\mu_{2n})\}_{n=1}^{\infty}$. Let $X'\subset X$ be the set of all $x\in X$ such that

$$\begin{aligned}\mathbf{E}[f|\mathbf{F}^2](x) &= \lim_{n\to\infty}\frac{1}{|S_{2n}(e)|}\sum_{g\in S_{2n}(e)} f(g^{-1}x)\\ &= \lim_{n\to\infty}\mathbf{A}'_\psi[f|\mathcal{R}_n](x)\,,\end{aligned}$$

$$\left(\mathbf{E}([|f|^{p'}|\mathbf{F}^2](x)\right)^{1/p'} = \lim_{n\to\infty}\left(\frac{1}{|S_{2n}(e)|}\sum_{g\in S_{2n}(e)}\left|f(g^{-1}x)\right|^{p'}\right)^{1/p'}.$$

By Proposition 5.4 and Corollary 5.8, $\lambda(X')=1$. For $x\in X'$ and $n>0$, let $f_{x,2n}\in\ell^{p'}(\mathbf{F})$ be the function $f_{x,2n}(g):=f(g^{-1}x)$ if $g\in S_{2n}(e)$ and $f_{x,2n}(g):=0$ otherwise. By Lemma 5.7 and Hölder's inequality for functions on $\mathbf{F}$,

$$\begin{aligned}\left|\mu_{2n}(f)(x)-\mathbf{A}'_\psi[f|\mathcal{R}_n](x)\right| &= \Big|\sum_{g\in S_{2n}(e)} f(g^{-1}x)\big(\mu_{2n}(g)-\eta^\psi_{2n}(g)\big)\Big|\\ &\le \|f_{x,2n}\|_{\ell^{p'}(\mathbf{F})}\|\mu_{2n}-\eta^\psi_{2n}\|_{\ell^q(\mathbf{F})}\,.\end{aligned}$$

Recall that $\pi_\partial\colon \ell^1(\mathbf{F})\to L^1(\partial\mathbf{F},\nu)$ is defined by $\pi_\partial(\delta_g)=\nu(O(g))^{-1}\chi_{O(g)}=|S_{2n}(e)|\chi_{O(g)}$ if $|g|=2n$. Since $O(g)$, $g\in S_{2n}$ form a partition of $\partial\mathbf{F}$, clearly

$$\left|\pi_\partial(\mu_{2n})(\xi)-\pi_\partial(\eta^\psi_{2n})(\xi)\right|^q=\sum_{g\in S_{2n}}\left|\mu_{2n}(g)-\eta^\psi_{2n}(g)\right|^q\nu(O(g))^{-q}\chi_{O(g)}(\xi)\,.$$

It now follows that

$$\begin{aligned}\|\mu_{2n}-\eta^\psi_{2n}\|_{\ell^q(\mathbf{F})} &= \Big(\sum_{g\in S_{2n}(e)}|\mu_{2n}(g)-\eta^\psi_{2n}(g)|^q\Big)^{1/q}\\ &= \left(\sum_{g\in S_{2n}(e)}\frac{1}{\nu(O(g))}\int_{O(g)}\left|\mu_{2n}(g)-\eta^\psi_{2n}(g)\right|^q\chi_{O(g)}(\xi)d\nu(\xi)\right)^{1/q}\\ &= \Big(\sum_{g\in S_{2n}(e)}\nu(O(g))^{q-1}\int_{O(g)}\left|\pi_\partial(\mu_{2n})(\xi)-\pi_\partial(\eta^\psi_{2n})(\xi)\right|^q d\nu(\xi)\Big)^{1/q}\\ &= |S_{2n}(e)|^{-1/p'}\|\pi_\partial(\mu_{2n})-\pi_\partial(\eta^\psi_{2n})\|_{L^q(\partial\mathbf{F},\nu)}\,.\end{aligned}$$

Combining this with the previous inequality, we conclude

$$\left|\mu_{2n}(f)(x) - \mathbf{A}'_{\psi}[f|\mathcal{R}_n](x)\right|$$
$$\leq |S_{2n}(e)|^{-1/p'} \|f_{x,2n}\|_{\ell^{p'}(\mathbf{F})} \|\pi_{\partial}(\mu_{2n}) - \pi_{\partial}(\eta^{\psi}_{2n})\|_{L^q(\partial\mathbf{F},\nu)} .$$

The definition of X' implies $|S_{2n}(e)|^{-1/p'} \|f_{x,2n}\|_{\ell^{p'}(\mathbf{F})}$ tends to

$$\mathbf{E}[|f|^{p'}|\mathbf{F}^2](x)^{1/p'}$$

as $n \to \infty$. Since by assumption $\pi_{\partial}(\mu_{2n})$ converges to ψ in $L^q(\partial\mathbf{F})$, Lemma 5.9 implies that $\|\pi_{\partial}(\mu_{2n}) - \pi_{\partial}(\eta^{\psi}_{2n})\|_{L^q(\partial\mathbf{F},\nu)}$ tends to zero as $n \to \infty$. So

$$\lim_{n\to\infty} \left|\mu_{2n}(f)(x) - \mathbf{A}'_{\psi}[f|\mathcal{R}_n](x)\right| = 0 .$$

The definition of X' now implies

$$\lim_{n\to\infty} \mu_{2n}(f)(x) = \mathbf{E}[f|\mathbf{F}^2](x) .$$

This proves the pointwise result if $1 < q < \infty$.

As to the case $q = \infty$, uniform convergence of $\pi_{\partial}(\mu_{2n})$ implies that the limit function ψ is continuous on the boundary. Therefore the second part of Lemma 5.9 gives the uniform convergence of $\pi_{\partial}(\eta^{\psi}_{2n})$ to ψ, and thus also the convergence of $\|\pi_{\partial}(\mu_{2n}) - \pi_{\partial}(\eta^{\psi}_{2n})\|_{L^{\infty}(\partial\mathbf{F},\nu)}$ to zero. Corollary 5.8 gives the convergence of $|S_{2n}(e)|^{-1}\|f_{x,2n}\|_{\ell^1(\mathbf{F})}$ to $\mathbf{E}[|f||\mathbf{F}^2](x)$ if $f \in L\log^+ L(X,\lambda)$. Using these two facts the same arguments used above establish the desired result also in the case when $p = p' = 1$ and $q = \infty$ provided $f \in L\log^+ L(X,\lambda)$.

Finally, we note that the fact that $\mu_{2n}(f)$ converges to $\mathbf{E}[f|\mathbf{F}^2]$ in L^p-norm (if $p > 1$) follows from the pointwise result by a standard argument (e.g., see the end of the proof of Theorem 2.1). □

REFERENCES

[AL05] AARONSON, J. and M. LEMAŃCZYK. Exactness of Rokhlin endomorphisms and weak mixing of Poisson boundaries. In: *Algebraic and Topological Dynamics*, 77–87. Contemp. Math. *385*, Amer. Math. Soc., Providence, RI, 2005.

[Bo10] BOWEN, L., Invariant measures on the space of horofunctions of a word hyperbolic group. *Ergodic Theory Dynam. Systems 30* (2010), 97–129.

[BN1] BOWEN, L. and A. NEVO. Pointwise ergodic theorems beyond amenable groups. *Ergodic Theory Dynam. Systems 33* (2013), 777–820.

[BN2] BOWEN, L. and A. NEVO. *A horospherical ratio ergodic theorem for actions of free groups*. Preprint arXiv: 1207.3569 (2012); to appear in *Groups, Geometry and Dynamics*.

[BN3] BOWEN, L. and A. NEVO. von-Neumann and Birkhoff ergodic theorems for negatively curved groups. Preprint arXiv: 1303.4109 (2013).

[BN4] BOWEN, L. and A. NEVO. Amenable equivalence relations and the construction of ergodic averages for group actions. Preprint arXiv: 1303.4426 (2013).

[BN5] BOWEN, L. and A. NEVO. Pointwise ergodic theorems for Markov groups. In preparation.

[Bu02] BUFETOV, A. I. Convergence of spherical averages for actions of free groups. *Ann. of Math. (2) 155* (2002), 929–944.

[BKK11] BUFETOV, A. I., M. KHRISTOFOROV and A. KLIMENKO. Cesaro convergence of spherical averages for measure-preserving actions of Markov semigroups and groups. Preprint arXiv: 1101.5459 (2011).

[BK] BUFETOV, A. and A. KLIMENKO. On Markov operators and ergodic theorems for group actions. *European J. Combin. 33* (2012), 1427–1443.

[BS10] BUFETOV, A. I. and C. SERIES. A pointwise ergodic theorem for Fuchsian groups. *Math. Proc. Cambridge Philos. Soc. 151* (2011), 145–159.

[DS] DUNFORD, N. and J. T. SCHWARTZ. *Linear Operators, Part I*. Reprint of the 1958 original. Wiley Classics Library. A Wiley-Interscience Publication. John Wiley & Sons, Inc., New York, 1988.

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

[FN98] FUJIWARA, K. and A. NEVO. Maximal and pointwise ergodic theorems for word-hyperbolic groups. *Ergodic Theory Dynam. Systems 18* (1998), 843–858.

[GN10] GORODNIK, A. and A. NEVO. *The Ergodic Theory of Lattice Subgroups*. Annals of Mathematics Studies *172*. Princeton University Press, Princeton, NJ, 2010.

[Gr99] GRIGORCHUK, R. I. Ergodic theorems for the actions of a free group and a free semigroup. *Mat. Zametki 65* (1999), 779–783; translation in *Math. Notes 65* (1999), 654–657.

[Ka00] KAIMANOVICH, V. A. The Poisson formula for groups with hyperbolic properties. *Ann. of Math. (2) 152* (2000), 659–692.

[Ka03] —— Double ergodicity of the Poisson boundary and applications to bounded cohomology. *Geom. Funct. Anal. 13* (2003), 852–861.

[Li01] LINDENSTRAUSS, E. Pointwise theorems for amenable groups. *Invent. Math. 146* (2001), 259–295.

[Ne94] NEVO, A. Harmonic analysis and pointwise ergodic theorems for non-commuting transformations. *J. Amer. Math. Soc. 7* (1994), 875–902.

[Ne05] —— Pointwise ergodic theorems for actions of groups. In: *Handbook of Dynamical Systems*, vol. 1B, Eds. B. Hasselblatt and A. Katok, 871–982. Elsevier B. V., Amsterdam, 2006.

[NS94] NEVO, A. and E. M. STEIN. *A generalization of Birkhoff's pointwise ergodic theorem*. *Acta Math. 173* (1994), 135–154.

[PS] POLLICOTT, M. and R. SHARP. Ergodic theorems for actions of hyperbolic groups. *Proc. Amer. Math. Soc. 141* (2013), 1749–1757.

[Rot62] ROTA, G.-C. An "alternierende Verfahren" for general positive operators. *Bull. Amer. Math. Soc. 68* (1962), 95–102.

[SW71] STEIN, E. M. and G. WEISS. *Introduction to Fourier Analysis on Euclidean Spaces.* Princeton Mathematical Series *32*. Princeton University Press, Princeton, N.J., 1971.

[W39] WIENER, N. The ergodic theorem. *Duke Math. J. 5* (1939), 1–18.

(Reçu le 8 mars 2012)

Lewis Bowen

The University of Texas at Austin
Department of Mathematics
1 University Station C1200
Austin, TX 78712-0257
U. S. A.
e-mail: lpbowen@math.utexas.edu

Amos Nevo

Mathematics Department
Technion - Israel Institute of Technology
Haifa, 32000
Israel
e-mail: anevo@tx.technion.ac.il