# Ergodic Theorems for Spherical Averages of Free Group Actions

Daniel Raban

## Contents

# 1 Introduction: Ergodic Theorems for Measure-Preserving Group Actions

The classical ergodic theorem concerns the convergence of "time-averaged" data in a measure-preserving system to a "spatial average" of the data:

**Theorem 1.1** (Pointwise + $L^p$ ergodic theorem)**.** *Fix $p \in [1, \infty)$, and let $(X, \mathcal{B}, \mu, T)$ be an ergodic measure-preserving system. If $f \in L^p$, then*

$$\frac{1}{N} \sum_{n=0}^{N-1} f \circ T^n \to \int f \, d\mu$$

*pointwise and in $L^p$ as $N \to \infty$.*

In this setting, $N$ is usually intuited as indexing time, and it encodes the information of a repeated measure-preserving action by $T$. If there are multiple measure-preserving actions by $T_1, \dots, T_d$, the situation becomes more complex. First, we will assume $T_1, \dots, T_d$ are **commuting**, i.e. $T_i \circ T_j = T_j \circ T_i$. Then the action of $T_1, \dots, T_d$ on the probability space can be indexed by how many times each of the $T_i$ have acted; that is, the joint action of $T_1, \dots, T_d$ is indexed by $\mathbb{Z}^d$, and we write $T^{(n_1, \dots, n_d)} := T_1^{n_1} \circ \cdots \circ T_d^{n_d}$. In this case, the ergodic theorem holds as usual:

**Theorem 1.2** ($L^p$ ergodic theorem for $\mathbb{Z}^d$ actions)**.** *Fix $p \in [1, \infty)$, let $(X, \mathcal{B}, \mu)$ be an probability space, and let $T_1, \dots, T_d$ be commuting, measure-preserving maps on $X$. Assume $X$ is ergodic with respect to $T_1, \dots, T_d$, i.e. if $T^{-i}E = E$ for all $i$, then $\mu(E) = 0$ or 1.If $f \in L^p$, then*

$$\frac{1}{|B_N|} \sum_{n \in B_N} f \circ T^n \xrightarrow{L^p} \int f \, d\mu$$

*as $N \to \infty$, where $(B_N)_N$ is an increasing sequence of boxes in $\mathbb{Z}_+^d$ converging to $\mathbb{Z}_+^d$.*

However, if $T_1, \dots, T_d$ are not commuting, the situation becomes more complex. The action, rather than being indexed by $\mathbb{Z}^d$, is indexed by some group with $d$ generators. To study general measure-preserving group actions, we move our attention to $\mathbb{F}_d$, the free group on $d$ generators. This corresponds to the situation where the actions of $T_1, \dots, T_d$ have no relations between them.

Analysis of ergodic free group actions is more difficult because the free group $\mathbb{F}_d$ is not *amenable*. Instead of discussing limits of averages over boxes, we will discuss *spherical averages*, which are averages over all words in $\mathbb{F}_d$ of length $n$.

## 2 The Calderón Transference Principle: Ergodic Theorems for Actions by Amenable Groups

To see why the situation for free group actions is difficult, we must first see what goes well in the case of $\mathbb{Z}^d$ and actions like it. The classical $L^1$ and pointwise ergodic theorems are usually proven using maximal inequalities. In the case of $\mathbb{Z}^d$, we can establish maximal inequalities on $L^p(X)$ using weaker maximal inequalities on $\ell^p(\mathbb{Z}^d)$. Here is the tool that allows us to establish such maximal inequalities. We will state the result for more general types of averages than just uniform averages on boxes; we state it for averages

$$A_N f := \sum_{j \in \mathbb{Z}^d} a_{N,j} f \circ T^j.$$

**Theorem 2.1** (Calderón transference principle)**.** *Let $(X, \mathcal{B}, \mu)$ be a probability space, and let $T_1, \ldots, T_d$ be commuting measure-preserving maps on $X$. Let $(a_{N,j})_{n \in \mathbb{N}, j \in \mathbb{Z}^d}$ be such that $\sum_{j \in \mathbb{Z}^d} |a_{N,j}| < \infty$ for all $N$. If*

$$\left\| \sup_{N \in \mathbb{N}} \left| \sum_{j \in \mathbb{Z}^d} a_{N,j} \psi(k-j) \right| \right\|_{\ell^p(k)} \leq C \|\psi\|_{\ell^p(\mathbb{Z}^d)},$$

*for all $\psi \in \ell^p(\mathbb{Z}^d)$, then*

$$\left\| \sup_{N \in \mathbb{N}} \left| \sum_{j \in \mathbb{Z}^d} a_{N,j} f(T^{-j} x) \right| \right\|_{L^p(x)} \leq C \|f\|_{L^p(X)}$$

*for all $f \in L^p(X)$.*

The following proof follows section 1.5 of [LaV10].

*Proof.* Step 1 (Time truncation): It suffices to show a bound

$$\left\| \sup_{1 \leq N \leq J} \left| \sum_{j \in \mathbb{Z}^d} a_{N,j} f(T^{-j} x) \right| \right\|_{L^p(x)} \leq C \|f\|_{L^p(X)}$$

which holds independent of $J$, so fix such a $J \in \mathbb{N}$.

Step 2 (Finite support of averages): We now claim that it suffices to assume that the averages are contained in some box; in particular, we claim that it suffices to assume that there exists a box $B$ such that $a_{N,j} = 0$ for $j \notin B$ and $1 \leq N \leq J$. Indeed, for any $\varepsilon > 0$, we can let $B_N$ be a box in $\mathbb{Z}^d$ such that $\sum_{j \notin B_N} |a_{N,j}| < \varepsilon 2^{-N}$.

Then letting $B = \bigcup_{N=1}^J B_N$, the additional error in the hypothesis and conclusion become $\sum_{N=1}^J \varepsilon 2^{-N} \leq \varepsilon$. Letting $\varepsilon \downarrow 0$, we recover the original result.

Step 3 (Transference): Given $f \in L^p(X)$ and $x \in X$, define

$$\psi(k) := \begin{cases} f(T^k x) & k \in K + B \\ 0 & \text{otherwise,} \end{cases}$$

where $K \subseteq \mathbb{Z}^d$ should be thought of as a large box with size we will send to $\infty$. Then for $k \in K$, $\sum_{j \in B} a_{N,j} f(T^{k-j}x) = \sum_{j \in B} a_{N,j} \psi(k-j)$. If we take the sup over $N$ and sum the $p$-th powers of both sides over $k$, we get

$$\begin{aligned} \sum_{k \in K} \sup_{1 \leq N \leq J} \left| \sum_{j \in \mathbb{Z}^d} a_{N,j} f(T^{k-j}x) \right|^p &= \sum_{k \in K} \sup_{1 \leq N \leq J} \left| \sum_{j \in \mathbb{Z}^d} a_{N,j} \psi(k-j) \right|^p \\ &\leq \left\| \sup_{1 \leq N \leq J} \left| \sum_{j \in \mathbb{Z}^d} a_{N,j} \psi(k-j) \right| \right\|_{\ell^p(k)}^p \\ &\leq C^p \|\psi\|_{\ell^p(\mathbb{Z}^d)}^p \\ &= C^p \sum_{k \in K+B} |f(T^k x)|^p. \end{aligned}$$

Now integrate over $x$ to get

$$\begin{aligned} \sum_{k \in K} \left\| \sup_{1 \leq N \leq J} \left| \sum_{j \in \mathbb{Z}^d} a_{N,j} f \circ T^{k-j} \right| \right\|_{L^p(X)}^p &\leq C^p \sum_{k \in K+B} \|f \circ T^k\|_{L^p(X)}^p \\ &= C^p |K+B| \cdot \|f\|_{L^p(X)}^p, \end{aligned}$$

using the $T_i$-invariance of $\mu$ for each $i$. Using the $T_i$-invariance on the left hand side of this inequality makes the summands equal to $\| \sup_{1 \leq N \leq J} | \sum_{j \in \mathbb{Z}^d} a_{N,j} f \circ T^{-j} | \|_{L^p(X)}^p$, which are independent of $k$. In total, we get

$$\left\| \sup_{1 \leq N \leq J} \left| \sum_{j \in \mathbb{Z}^d} a_{N,j} f \circ T^{-j} \right| \right\|_{L^p(X)}^p \leq C^p \frac{|K+B|}{|K|} \|f\|_{L^p(X)}^p.$$

Letting $K \uparrow \mathbb{Z}^d$ and taking $p$-th roots, we get

$$\left\| \sup_{1 \leq N \leq J} \left| \sum_{j \in \mathbb{Z}^d} a_{N,j} f \circ T^{-j} \right| \right\|_{L^p(X)}^p \leq C^p \frac{|K+B|}{|K|} \|f\|_{L^p(X)} \leq C \|f\|_{L^p(X)},$$

as desired. $\square$

This argument hinges on the ability to cover $\mathbb{Z}^d$ with a sequence of increasing finite boxes $K$ in a nice enough way such as to make the truncation arguments work. If we index our measure-preserving action by a more general group, this approximation is often but not always possible. In particular, this argument applies to *amenable groups*.

**Definition 2.1.** An **amenable group** $G$ is a locally compact, countable, Hausdorff topological group that admits a **Følner sequence**, an increasing sequence $F_1, F_2, \ldots$ of finite subsets of $G$ such that for every $g \in G$,

$$\lim_{n\to\infty} \frac{|gF_k \triangle F_k|}{|F_k|} = 0.$$

Amenable groups are the ones that can be approximated well by finite sets. For example, a Følner sequence for $\mathbb{Z}^d$ are the boxes $B_n = \{(x_1, \ldots, x_d) \in \mathbb{Z}^d : |x_1|, \ldots, |x_d| \leq n\}$. The argument in the Calderón transferrence principle applies to any amenable group $G$ if you replace the boxes $K \uparrow \mathbb{Z}^d$ by a Følner sequence for $G$. However, this argument does not hold for $\mathbb{F}_d$.

**Theorem 2.2.** *For $d \geq 2$, $\mathbb{F}_d$ is not amenable.*

The proof of this theorem is based on what is known as a **paradoxical decomposition** of $\mathbb{F}_d$. In particular, $\mathbb{F}_d = \{e\} \sqcup A_{g_1} \sqcup A_{g_1^{-1}} \sqcup \cdots \sqcup A_{g_d} \sqcup A_{g_d^{-1}}$, where $A_{g_i}$ is the set of all words in $\mathbb{F}_d$ starting with the generator $g_i$. Each $A_{g_i}$ looks like $\mathbb{F}_d$, in the sense that $g_i^{-1} A_{g_i} = \mathbb{F}_d$. Because of this structure, it is difficult for $g_i F$ to be very distinct from $F$ because applying $g_i$ to $F$ pushes it completely into the territory of $A_{g_i}$. So to avoid too much overlap, we would need $F$ to be mostly contained in one of the $A_{g_1}$; but then the same problem arises again. Here is a proof to formally express this intuition:

*Proof.* Suppose that for $\varepsilon > 0$, we have a finite subset $F \subseteq \mathbb{F}_d$ such that

$$\frac{|gF \triangle F|}{|F|} < \varepsilon$$

for $g = g_1, \ldots, g_d, g_1^{-1}, \ldots, g_d^{-1}$. By inclusion-exclusion,

$$|g_1 F \triangle F| = |g_1 F| + |F| - 2|g_1 F \cap F| = 2|F| - 2|g_1 F \cap F|,$$

so we can rearrange to get

$$\frac{|g_1 F \cap F|}{|F|} > 1 - \frac{\varepsilon}{2}.$$

Since $g_1 F \subseteq A_{g_1}$, we get

$$\frac{|A_{g_1} \cap F|}{|F|} > 1 - \frac{\varepsilon}{2}.$$

If we apply this same argument to the other generators and add the bounds together, we get

$$\frac{|F \setminus \{e\}|}{|F|} > d\left(1 - \frac{\varepsilon}{2}\right).$$

The left hand side is $< 1$, while for $\varepsilon \leq 2(1 - 1/d)$, the right hand side must be $\geq 1$. So we cannot have the Følner property for small $\varepsilon$. $\square$

# 3 The Nevo-Stein and Bufetov Ergodic Theorems for Spherical Average Operators

## 3.1 Markov operators

Instead of averaging over an increasing Følner sequence, as in the case of amenable groups, for free groups, it is more natural (and more fruitful) to study uniform averages over words of length $n$. When $\mathbb{F}_d$ is viewed as a Cayley graph generated by the natural $d$ generators of the group, the (reduced) words of length $n$ are the elements at distance $n$ from the identity element, i.e. the spherical shell of radius $n$. Note that there are $2d(2d-1)^{n-1}$ words of length $n$, as we cannot pick $g_i^{-1}$ after picking $g_i$ in the word, lest we shorten the word.

Before we introduce an ergodic theorem in this case, we will switch our point of view slightly to a functional analytic perspective. Recall that given a measure-preserving $T : X \to X$, there is an associated **Koopman operator** $U_T : L^p(X) \to L^p(X)$ for each $p \in [1, \infty]$, given by $U_T(f) = f \circ T$. Through a minor overloading of notation, we will use $T_g$ to mean the Koopman operator $U_{T_g}$. In the language of Koopman operators, $L^p$ ergodic theorems assert that $A_N f$ converge for $f \in L^p$, where $A_N$ is an average of Koopman operators $T_g$. This is convergence of $A_N$ in the strong operator topology on $\mathcal{L}(L^p(X), L^p(X))$.

It will be useful for us to generalize this type of operator and consider general *Markov operators*.

**Definition 3.1.** Let $(X, \mathcal{B}_X, \mu)$ and $(Y, \mathcal{B}_Y, \nu)$ be measure spaces. An operator $U : L^1(Y) \to L^1(X)$ is called a **Markov operator** if

(i) (Positivity) $Uf \geq 0$ when $f \geq 0$,

(ii) (Preserves identity) $U\mathbb{1}_Y = \mathbb{1}_X$,

(iii) (Preserves integration) $\int_X Uf \, d\mu = \int_Y f \, d\nu$ for all $f \in L^1(Y)$.

**Example 3.1.** An important example of Markov operators are the **transition operators** of Markov chains. In particular, if $(Z_n)_{n \in \mathbb{Z}}$ is a Markov chain with state space $X$, then the map $P : L^1(X) \to L^1(X)$ given by $Pf := \mathbb{E}[f(Z_1) \mid Z_0]$ is a Markov operator. When the $X$ is finite, this can be thought of as the transition matrix of the chain, which is a linear map $P : \mathbb{R}^{|X|} \to \mathbb{R}^{|X|}$. Note that by the Markov property, $Pf = \mathbb{E}[f(Z_{n+1}) \mid Z_n]$ for any $n \in \mathbb{Z}$.

We will need the following theorem concerning Markov operators:

**Lemma 3.1** (Rota)**.** *Fix $p \in (1, \infty)$, and let $P$ be a Markov operator on $L^1(X)$. For every $f \in L^p(X)$, the sequence $(P^*)^n P^n f$ converges almost everywhere and in $L^p$ as $n \to \infty$.*

To prove this, recall the backwards martingale convergence theorem:

**Lemma 3.2** (Backwards martingale convergence)**.** *Let $(f_n)_{n \le 0}$ be a backwards martingale adapted to the decreasing filtration $\mathcal{F}_0 \supseteq \mathcal{F}_{-1} \supseteq \cdots$. Then the limit $f_{-\infty} := \lim_{n \to -\infty} f_n$ exists a.s. and in $L^1$. Moreover, if $f_0 \in L^p$, the convergence occurs in $L^p$, as well.*

*Proof.* See section 4.7 of [Dur19]. □

*Proof.* Consider the sequence space $X^{\mathbb{Z}}$, and let $\widetilde{\mu}$ be a shift-invariant measure given on $X^{\mathbb{Z}}$ so that $(X^{\mathbb{Z}}, \widetilde{\mu})$ is the space of trajectories of a Markov chain with transition operator $P$ and stationary distribution $\mu$. With this setup, $Pf(x_n) = \mathbb{E}[f(x_{n+1}) \mid \mathcal{F}_n]$, where $\mathcal{F}_n$ is the $\sigma$-algebra on the $n$-th coordinate. The adjoint $P^*$ (that is, the time reversed transition operator) likewise satisfies $P^* f(x_n) = \mathbb{E}[f(x_{n-1}) \mid \mathcal{F}_n]$.

For $k, \ell \in \{-\infty\} \cup \mathbb{Z} \cup \{\infty\}$, let $\mathcal{F}_k^\ell = \sigma(\mathcal{F}_k, \mathcal{F}_{k+1}, \cdots, \mathcal{F}_\ell)$ be the $\sigma$-algebra generated by the coordinates $k, k+1, \ldots, \ell$. Given $f \in L^1(X)$, define the map $F \in L^1(X^{\mathbb{Z}}, \widetilde{\mu})$ by $F(x) = f(x_0)$. We claim that

$$\mathbb{E}[F(x) \mid \mathcal{F}_{-\infty}^{-n}] = P^n f(x_{-n}), \qquad \mathbb{E}[\mathbb{E}[F(x) \mid \mathcal{F}_{-\infty}^{-n}] \mid \mathcal{F}_0] = (P^*)^n P^n f(x_0).$$

We prove the first claim by induction on $n$: If $n = 0$, $\mathbb{E}[F(x) \mid \mathcal{F}_{-\infty}^0] = \mathbb{E}[f(x_0) \mid \mathcal{F}_{-\infty}^0] = f(x_0)$. Now, supposing the claim is true for $n$, if we let $\pi_0$ be the projection onto the 0th coordinate,

$$\begin{aligned} P^{n+1} f(x_{-(n+1)}) &= P(P^n f((\sigma^{-1} x)_{-n})) \\ &= P(\mathbb{E}[F(x) \mid \mathcal{F}_{-\infty}^{-n}] \circ \sigma^{-1}) \\ &= \mathbb{E}[\mathbb{E}[F(x) \mid \mathcal{F}_{-\infty}^{-n}] \mid \mathcal{F}_{-\infty}^{-(n+1)}] \end{aligned}$$

Using the tower property of conditional expectation,

$$= \mathbb{E}[F(x) \mid \mathcal{F}_{-\infty}^{-(n+1)}].$$

To prove the second claim, we need to show $\mathbb{E}[P^n f(x_{-n}) \mid \mathcal{F}_0] = (P^*)^n P^n f(x_0)$. It suffices to show that $\mathbb{E}[h(x_{-n}) \mid \mathcal{F}_0] = (P^*)^n h(x_0)$ for any $h$. We proceed again by induction. When $n = 0$, both sides are equal to $h(x_0)$. Supposing this claim is true for $n$, we have

$$(P^*)^{n+1} h(x_0) = P^* \, \mathbb{E}[h(x_{-n}) \mid \mathcal{F}_0](x_0)$$

$$
\begin{aligned}
&= \mathbb{E}[\mathbb{E}[h(x_{-n+1}) \mid \mathcal{F}_0] \mid \mathcal{F}_0] \\
&= \mathbb{E}[h(x_{-(n+1)}) \mid \mathcal{F}_0],
\end{aligned}
$$

completing the induction.

Now, applying the backwards martingale convergence theorem, we get that the limit $\lim_{n\to\infty}(P^*)^n P^n f(x_0) = \lim_{n\to\infty} \mathbb{E}[\mathbb{E}[F(x) \mid \mathcal{F}_{-\infty}^{-n}] \mid \mathcal{F}_0]$ exists a.s. and in $L^p$. $\square$

## 3.2 The Nevo-Stein theorem for spherical averages

Here is an ergodic theorem for this case of spherical averages:

**Theorem 3.1** (Nevo-Stein, 1994). *Fix $p \in (1, \infty)$, and let $(X, \mathcal{B}, \mu)$ be a probability space, Define the uniform spherical average operators*

$$
s_n := \frac{1}{2d(2d-1)^{n-1}} \sum_{|w|=n} T_w,
$$

*where if $w = g_1 \cdots g_n$, then $T_w$ denotes $T_{g_1} \circ \cdots \circ T_{g_n}$. If $f \in L^p(X)$, then $s_{2n} f$ converges pointwise and in $L^p$.*

**Remark 3.1.** The theorem also works if we replace $s_{2n}$ by $s_{2n+1}$. The reason we can't just look at $s_n$ is that if $f$ is an eigenfunction of all the $T_{g_i}$ of eigenvalue $-1$, then $s_n f = (-1)^n f$. However, the $2n$ case works fine, as we will see.

To prove this, we will use an approach from [Buf02]. The idea is to construct a Markov operator $P$ that represents the spherical averaging, relate $P^{2n}$ to $(P^*)^n P^n$, and then apply Rota's theorem.

*Proof.* To construct a Markov operator which represents the spherical averaging, we must ensure that the operator takes into account where the walk came from (so as to not follow up $T_{g_i}$ with $T_{g_i^{-1}}$ and reduce the length of the word). To this end, we essentially lift the random walk on the Cayley graph to a random walk on the until tangent bundle of the Cayley graph. In particular, if we let $A$ be the set of generators of $\mathbb{F}_d$ (and their inverses) and let $Y = X \times A$, then we can define the Markov operator $P : L^1(Y) \to L^1(Y)$ given by

$$
PF(x, i) = \frac{1}{2d-1} \sum_{j \neq i^{-1}} F(T_i x, j).
$$

In words, $Pf$ returns the average value of $f$ after taking one random step along the Cayley graph of $\mathbb{F}_d$, making sure not to backtrack. The second index queues up the next generator to be used. It follows that upon $n$-fold iteration, we have

$$P^n F(x,i) = \frac{1}{(2d-1)^n} \sum_{w=g_n\cdots g_1 : g_1 \neq i^{-1}} F(T_{g_{n-1}} \cdots T_{g_1} T_i x, g_{n-1}),$$

If $f \in L^1(X)$, then defining $F(x,i) = f(x)$,

$$s_n f(x) = \frac{1}{2d} \sum_{i \in A} P^n F(x,i).$$

So to show that $s_{2n} f$ converges, it suffices to show that $P^{2n}F$ converges.

We will now relate $P^{2n}$ to $(P^*)^n P^n$ and show that the latter converges when applied to $F$. First, we figure out what $P^*$ looks like: Observe that if $\eta = \mu \times (\frac{1}{2d} \sum_{i \in A} \delta_i)$ denotes the measure on $Y$, then

$$\begin{aligned}
\int_Y P^* F(x,i) H(x,i)\, d\eta &= \int_Y F(x,i) PH(x,i)\, d\eta \\
&= \frac{1}{2d-1} \sum_{j \neq i^{-1}} \int_Y F(x,i) H(T_i x, j)\, d\eta
\end{aligned}$$

Leveraging the $T_i \times \mathrm{id}_A$-invariance of $\eta$,

$$= \frac{1}{2d-1} \sum_{j \neq i^{-1}} \int_Y F(T_{i^{-1}} x, i) H(x, j)\, d\eta$$

Now leveraging the invariance of $\eta$ under transpositions of $i, j$,

$$\begin{aligned}
&= \frac{1}{2d-1} \sum_{j \neq i^{-1}} \int_Y F(T_{j^{-1}} x, j) H(x, i)\, d\eta \\
&= \int_Y \left( \frac{1}{2d-1} \sum_{j \neq i^{-1}} F(T_{j^{-1}} x, j) \right) H(x,i)\, d\eta.
\end{aligned}$$

So we get that

$$P^* F(x,i) = \frac{1}{2d-1} \sum_{j \neq i^{-1}} F(T_{-j} x, j).$$

for $F \in L^2(Y)$, and this extends to $F \in L^1$ by the density of $L^2(Y)$ in $L^1(Y)$. In other words, $P^*$ averages $f$ with respect to the random walk run in reverse.

Let $U$ be the unitary operator $UF(x,i) = F(T_i x, i^{-1})$ which walks in the direction of $i$ and then queues up $i^{-1}$ (so that $U^2 = I$). First observe that $P = UP^*U$:

$$\begin{aligned} UP^*UF(x,i) &= UP^*F(T_i x, i^{-1}) \\ &= \frac{1}{2d-1}\sum_{j\neq i} UF(T_{j^{-1}}T_i x, j) \\ &= \frac{1}{2d-1}\sum_{j\neq i} UF(T_i x, j^{-1}) \end{aligned}$$

Reindexing with $j$ instead of $j^{-1}$,

$$= PF(x,i).$$

We now claim that

$$(P^*)^n P^n = \frac{2d-2}{2d-1} UP^{2n-1} + \frac{1}{2d-1}(P^*)^{n-1}P^{n-1}.$$

Given this claim, we will be done because Rota's theorem implies that all terms except the $UP^{2n-1}$ term converge. So $P^{2n} = PP^{2n-1}$ converges.

We prove the claim by induction on $n$. For $n = 1$, we have

$$\begin{aligned} PP^*F(x,i) &= \frac{1}{2d-1}\sum_{j\neq i^{-1}} PF(T_{j^{-1}}x, j) \\ &= \frac{1}{(2d-1)^2}\sum_{j\neq i^{-1}}\sum_{k\neq j^{-1}} F(x,k) \end{aligned}$$

The term where $k = i$ appears $2d-1$ times in the sum (once for each $j$), while the terms where $k \neq i$ each appear $2d-2$ times (once for each $j$ except the inverse of the desired $k$).

$$\begin{aligned} &= \frac{2d-2}{(2d-1)^2}\sum_{k\neq i} F(x,k) + \frac{1}{2d-1}F(x,i) \\ &= \frac{2d-2}{2d-1} PF(T_i x, i^{-1}) + \frac{1}{2d-1}F(x,i) \\ &= \frac{2d-2}{2d-1} UPF(x,i) + \frac{1}{2d-1}F(x,i) \\ &= \left(\frac{2d-2}{2d-1}UP + \frac{1}{2d-1}I\right)F(x,i). \end{aligned}$$

Now, supposing the claim is true for $n$, observe that

$$\begin{aligned}(P^*)^{n+1}P^n &= P^*\left(\frac{2d-2}{2d-1}UP^{2n-1} + \frac{1}{2d-1}(P^*)^{n-1}P^{n-1}\right)P \\ &= \frac{2d-2}{2d-1}P^*UP^{2n+1-1} + \frac{1}{2d-1}(P^*)^{(n+1)-1}P^{(n+1)-1}\end{aligned}$$

Using $P = UP^*U \implies P^*U = UP$,

$$= \frac{2d-2}{2d-1}UP^{2(n+1)-1} + \frac{1}{2d-1}(P^*)^{(n+1)-1}P^{(n+1)-1},$$

proving the claim. □

## 3.3 Bufetov's theorem for non-uniform spherical averages

The paper [Buf02] strengthens this result in multiple ways. Instead of uniform spherical averages, Bufetov proved convergence for a larger class of spherical averaging operators. The spherical averages are constructed by essentially taking a (possibly weighted) random walk on the Cayley graph of $\mathbb{F}_d$, rather than just a uniform random walk.

**Theorem 3.2** (Bufetov, 2002)**.** *Let $\Pi = (p_{i,j})_{i,j\in A}$ be a stochastic $2d\times 2d$ matrix with a unique stationary distribution $(p_{-d}, \ldots, p_{-1}, p_1, \ldots, p_d)$ with all $p_i > 0$ satisfying the following two conditions:*

*(i) (Generates $\mathbb{F}_d$): $p_{i,j} = 0$ iff $i = j^{-1}$.*

*(ii) (Symmetry/detailed balance): For all $i, j \in A$,*

$$p_i = p_{-i}, \qquad p_{-i,-j} = \frac{p_j p_{j,i}}{p_i}.$$

*Define the spherical averaging operators*

$$s_n^\Pi = \sum_{|w|=n} p_{w_n} p_{w_n,w_{n-1}} p_{w_{n-1},w_{n-2}} \cdots p_{w_2,w_1} T_w.$$

*Then for any $f$ such that $\int_X |f(x)| \log^+ |f(x)|\, d\mu(x) < \infty$, $s_{2n}^\Pi f$ converges $\mu$-a.e. and in $L^1(X)$ to an function invariant under $T_w$ for all words $w$ of even length.*

**Remark 3.2.** The symmetry condition (ii) implies that the spherical average operators $s_{2n}^{\Pi}$ are self-adjoint. The theorem also still holds if we replace condition (i) with a slightly weaker condition, strict irreducibility of $\Pi$. $\Pi$ is called **strictly irreducible** if both $\Pi$ and $\Pi\Pi^{\top}$ are irreducible as Markov chain transition matrices.

To understand the meaning of this strict irreducibility condition, observe that

$$[\Pi\Pi^{\top}]_{i,j} = \sum_k p_{i,k} p_{j,k},$$

so if $[PP^{\top}]_{i,j} > 0$, then there is some $k$ such that $i$ and $j$ can lead to $k$ in 1 step. In this sense, strict irreducibility is a kind of high connectivity condition on the directed graph corresponding to the Markov chain with transition matrix $\Pi$. The idea is that if you have a property which is shared by all states in the Markov chain which lead to the same state $k$, then strict irreducibility allows you to extend this property to all states in the chain.

We will not provide the entire proof, since the overall argument follows the general strategy we have used to prove the Nevo-Stein theorem but with additional details. Here are some of the main differences in this general case:

The following lemma provides the scaffolding used to achieve $L^1$ convergence instead of $L^p$ convergence.

**Lemma 3.3.** *Let $P$ be a Markov operator on a probability space $(Y, \eta)$. Then the tail $\sigma$-algebra $\bigcap_{k=1}^{\infty} \mathcal{F}_k^{\infty}$ of the Markov process induced by $P$ is trivial if and only if for any $f \in L^1(Y)$, $(P^*)^n f \xrightarrow{L^1} \int_Y f \, d\eta$ as $n \to \infty$.*

So to prove $L^1$ convergence, the strategy becomes proving that the tail $\sigma$-algebra of the Markov operator $P$ on $L^1(Y)$ (with $Y = X \times A$) is trivial. In this situation, we are dealing with $P^{2n}$, which can be expressed in terms of $(P^*)^n P^n$ as above.

Here is how we get pointwise convergence. Let $\|f\|_{L \log L} := \int_X |f(x)| \log^+ |f(x)| \, d\mu(x)$, and let $L \log L := \{f : \|f\| < \infty\}$. The relationship between this $L \log L$ condition and $L^1$ convergence is given by the following lemma.

**Lemma 3.4.** *There exists a constant $C > 0$ such that for any $f \in L \log L$,*

$$\|\sup_n (P^*)^n P^n f\|_{L^1} \leq C \|f\|_{L \log L}.$$

Combining the pointwise convergence of $s_{2n}^{\Pi} f$ for $f \in L^2(X)$ (using the same arguments as the Nevo-Stein theorem) with the above maximal inequality and the density of $L^2(X)$ in $L \log L$ gives pointwise convergence of $s_{2n}^{\Pi} f$ for $f \in L \log L$.

### 3.4 Failure of convergence for spherical averages of $L^1$ functions

In the Nevo-Stein theorem, we showed that the spherical averages $s_{2n}f$ converge for $f \in L^p$ for any $p > 1$. Bufetov's theorem upgrades this result to apply to $f \in L \log L$, which contains $L^p$ for each $p > 1$. However, the question remains of whether the convergence occurs for all $f \in L^1(X)$, as it does in the classical ergodic theorems. Unfortunately, the convergence fails in general for $L^1$ functions.

**Theorem 3.3** (Tao, 2015)**.** *There exists a probability space $(X, \mathcal{B}, \mu)$ with a measure-preserving $\mathbb{F}_2$-action $(T_g)_{g\in\mathbb{F}_2}$ on $X$ and an $f \in L^1(X)$ such that $\sup_n |s_{2n}f(x)| = \infty$ for $\mu$-almost every $x \in X$. In particular, $s_{2n}f(x)$ does not converge for $\mu$-almost every $x \in X$.*

Tao's construction, while written for $\mathbb{F}_2$, generalizes to any $\mathbb{F}_d$ with $d \geq 2$.The construction is a modified version of a counterexample due to Ornstein, which exhibits the failure of the maximal ergodic theorem for $L^1$ functions. The idea of the construction is to make it so $s_{2n}f$ only becomes large for very large $n$, where the space $X$ is designed so that once the size of $s_{2n}f(x)$ becomes relevant, the dynamics have relegated $x$ to be in some portion of the space with small measure. This allows the $\|f\|_{L^1(X)} < \infty$, even when $\sup_n |s_{2n}f(x)| = \infty$ for $\mu$-a.e. $x$.

## References


[BK12] Alexander Bufetov and Alexey Klimenko. On markov operators and ergodic theorems for group actions. *European Journal of Combinatorics*, 33(7):1427–1443, 2012. Groups, Graphs, and Languages.

[Buf02] Alexander I Bufetov. Convergence of spherical averages for actions of free groups. *Annals of mathematics*, 155(3):929–944, 2002.

[Dur19] Rick Durrett. *Probability: theory and examples*, volume 49. Cambridge university press, 2019.

[EFHN15] Tanja Eisner, Bálint Farkas, Markus Haase, and Rainer Nagel. *Operator theoretic aspects of ergodic theory*, volume 272. Springer, 2015.

[EW13] Manfred Einsiedler and Thomas Ward. *Ergodic theory*. Springer, 2013.

[LaV10] Patrick Reilly LaVictoire. *Pointwise Ergodic Theorems for Nonconventional L ˆ 1 Averages*. PhD thesis, UC Berkeley, 2010.

[Tao15] Terence Tao. Failure of the pointwise and maximal ergodic theorems for the free group. In *Forum of Mathematics, Sigma*, volume 3. Cambridge University Press, 2015.